

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 05:31:50 ; Search time 2313.83 Seconds
(Without alignments)
16279.401 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atgctctcttcacgcgaac.....tgcgcgaagcaggtcgtgtga 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	100.0	1800	100.0	1800
2	100.0	1800	100.0	1800
3	100.0	1800	100.0	1800
4	100.0	1800	100.0	1800
5	100.0	1800	100.0	1800
6	100.0	1800	100.0	1800
7	100.0	1800	100.0	1800
8	100.0	1800	100.0	1800
9	100.0	1800	100.0	1800
10	100.0	1800	100.0	1800
11	100.0	1800	100.0	1800
12	100.0	1800	100.0	1800
13	100.0	1800	100.0	1800
14	100.0	1800	100.0	1800
15	100.0	1800	100.0	1800
16	100.0	1800	100.0	1800
17	100.0	1800	100.0	1800
18	100.0	1800	100.0	1800
19	100.0	1800	100.0	1800
20	100.0	1800	100.0	1800
21	100.0	1800	100.0	1800
22	100.0	1800	100.0	1800
23	100.0	1800	100.0	1800
24	100.0	1800	100.0	1800
25	100.0	1800	100.0	1800
26	100.0	1800	100.0	1800
27	100.0	1800	100.0	1800
28	100.0	1800	100.0	1800
29	100.0	1800	100.0	1800
30	100.0	1800	100.0	1800
31	100.0	1800	100.0	1800
32	100.0	1800	100.0	1800
33	100.0	1800	100.0	1800

1	1800	100.0	1800	6	AX148306
2	1800	100.0	3869	8	AB026549
3	1800	100.0	52232	8	AB028617
4	1798.4	99.9	2331	8	AY056255
5	862.8	47.9	95769	8	AC013430
6	824.6	45.8	1818	8	AX148316
7	824.6	45.8	2171	8	LENEOXANT
8	809.4	45.0	2164	8	STU276244
9	798.4	44.4	1839	6	AX148312
10	798.4	44.4	2432	6	AB030293
11	791.6	44.0	2398	8	AF190462
12	750.8	41.7	72058	8	AC074176
13	738.2	41.0	1752	6	AX148302
14	738.2	41.0	110102	8	ATF28J12
15	738.2	41.0	194143	8	ATCHRIV48
16	696.8	38.7	2153	8	AF224672
17	665	36.9	2310	8	AF224671
18	606.6	33.7	1815	6	AX148314
19	606.6	33.7	2498	8	ZMU95953
20	528.2	29.3	1734	6	AX148310
21	528.2	29.3	78921	8	AB028621
22	498.4	27.7	82697	8	ATT9A21
23	314.6	17.5	602	8	AF399655
24	177.4	9.9	1971	8	MD293765
25	169.6	9.4	1788	6	AX148304
26	169.6	9.4	2064	8	AY056789
27	169.6	9.4	96574	8	ATT18B16
28	169.6	9.4	199382	8	ATCHRIV50
29	169.2	9.4	1950	6	AX085456
30	155.8	8.7	1769	8	AY029525
31	143.6	8.0	122241	2	AP004129
32	139.8	7.8	1617	6	AX148308
33	139.8	7.8	1617	6	AX148318
34	139.8	7.8	1846	8	ATH5813
35	106.8	5.9	1967	8	AF224670
36	71.4	4.0	630	8	CSA416714
37	63.2	3.5	138541	2	OSJN00215
38	57.8	3.2	34869	1	AF324838
39	57.2	3.2	657	8	CSA416712
40	56.2	3.1	39428	1	AF322256
41	56	3.1	84730	2	AC074054
42	56	3.1	140864	2	AC074355
43	55.6	3.1	15226	1	AE006980
44	55.6	3.1	24680	1	MTY21C12
45	54.8	3.0	187126	2	AP004729

ALIGNMENTS

RESULT	1
AX148306	1800 bp
LOCUS	Sequence 5 from Patent EP116794.
DEFINITION	AX148306
ACCESSION	AX148306.1
VERSION	GI:14347193
KEYWORDS	
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS	Tsuchi, S., Kobayashi, M. and Shinozaki, K.
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene
JOURNAL	Patent: EP 1116794-A 5 18-JUL-2001;
FEATURES	Riken (JF)
SOURCE	Location/Qualifiers
1..1800	
/organism="Arabidopsis thaliana"	
/db_xref="taxon:3702"	
1..1800	
/note="unnamed protein product"	

```
/codon_start=1
/protein_id="CAC41194.1"
/db_xref="GI:14347194"
/translation="MASPTAAVSGRMIGVGNHOPPLSSOSSDLVSYCSLPMASRV
TRKINVSALHTPPALHPKROSSNSPALVYKPKKESSTKOMNLFORAAALDAABG
FLVSHKEHLPLKPADPSVQIAGNEAPVNEPVRNRLNVKLPDLSIKGVYRNGANP
LHEPVTHGHFEDGGMVAHAKVEHGSASVACRTQTNFVEROLGRVFPKALGELH
GHTGILARMLFYPARAAGIVDPAGTGVANAGLVFNGLRILMSBDDLPYOVLTPNG
DLKTVGRDFDGFESTMIAHPKDPESGELFALSYDVSKPYLKYPFSPDGRKSPD
VEIOLDPTMMHDAITENFVYVPDOVFEFLPMIRGSGSPVYDKNKVAREGLIDKY
AEDSSNIMKIDAPRCFCHLMANNEPRTDEVYVIGSMTPPDSIFNSDENLSVLS
EIRLNLTGSESTRPIISNEDQVNLLEGMNRMNLGKRTFAVLALAEPMFKVSGFA
KVDLTGVKRNHLIDSNRYGGEPLFLPEGEDEDEGYTLICEVHDEKTKWSLQVNAV
SLEVEATVKLSRPVPGFHFHTGICADDLAKQV"
```

BASE COUNT 458 a 464 c 439 g 439 t

ORIGIN

Query Match 100.0%; Score 1800; DB 6; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggctcttccaagcgaagcgcgtgttctggagagatggctggcgcaataact 60
DB 1 ATGGCTTCTTTCACGGCAACGGCTCGGTTCTGGAGATGGCTTGGCAATCATACT 60

QY 61 cagccgcacattacgtctctccaagctccgacttgagttatgtagctcctactatg 120
DB 61 CAGCGGCCATTATTCGTCTTCTAAAGCTCCGACTTGAGTTATGTAGCTCTTACCTTANG 120

QY 121 gccagtcgttacaagcgtlaagcgtcaatgttcaatctgcgttcaactcctcactct 180
DB 121 GCCACTCGGTACACGTAGCTCAATGTTTCATCGCTGCGCTTACACACTCTCCAGCTCTT 180

QY 181 cattccctaaagcaatcactaaactcccgccatgtgtgttaagcccaaaagcaaga 240
DB 181 CATTTCCTTAAAGCAATCATCAAACTCTCCGCAATGTTTAAAGCCCAAAAGCAAGA 240

QY 241 tccaacactaaacagatgaattgttccagagacggcgcgacgacgtttgacgcgacg 300
DB 241 TCCACACTAAACAGATGAATTTGTTCCAGNAGCGCGCGCGACGCTTGGACCGCGCG 300

QY 301 gaaggcttccctgtcagccacgagaagctacacccgcttccctaaacggcgtgactcctagt 360
DB 301 GAGGGTTCCTTGTACGACACAGAGACTACACCGCTTCTTAAACCGGCTGATCTAGT 360

QY 361 gttcagatccgcgcgaatatttgcgtgtgaatgaacagccgcctccggtatcttccg 420
DB 361 GTTCAGATCCGCCGAATTTTGTCTCCGGTGAATGAACACGCCCTCCGGGTATCTTCCG 420

QY 421 gtgctcggaaaactcccgatccatccaagaagtgtatgtgcgaacggaactaaccca 480
DB 421 GTGGTCGGAAAACTTCCGATTCATCAAGAAGTGTATGTCCGCAACGCACTTAACCCA 480

QY 481 cttaacgacgggtgacaggttacaactcttcgaacggaagacgtatgttcaagcgcgtc 540
DB 481 CTTTACGAGCCCGGTACAGGTCACCACTTCTTCGACGAGAGAGTATGTTTACCCCGTC 540

QY 541 aaatccgaacaggttgcagctagctagcttgcggtttactcaactcaaacgggtttgt 600
DB 541 AAATTCGAACAGGTTTCACTAGCTTACGCTTCCGGTTTACTCAAGCTTAACCGGTTTGT 600

QY 601 caggaacgtlcaattgtgtcgacggcttctcccaaaagccatcggtgagcttcaagcgccac 660
DB 601 CAGGAACGTCAATTGGGTGACCGGTTTCCCCAAAGCATGGGTGAGTTTACCGGCCAC 660

QY 661 accggtatgcccgaactatgtatcttaacgcagagctgcaagcggatagtcgaacgg 720
DB 661 ACCGATATTGCCGCACTATGCTATTCTACGCCAAGCTGACGCGGTATATGCAACCG 720

QY 721 gcaacggaacccggtgtagcttaacgccggttgcgtatcttaacggcggttatgtgt 780
DB 721 GCACACGGAACCGGTGTAGCTTAACGCCGGTTGTCTATTTCAAATGCGCGGTATTGGCT 780

QY 781 atctcggaagatgatttaacttaccacttcaacagttcagatcactcccatggagatttaaac 840
DB 781 ATCTCGGAAGATGATTTAACCTTACCAAGTTCAGATCACCCTCCAAATGGAGATTTAAC 840

QY 841 gtgtgtcgttgcatttgaatgaacatagaatccaaatgattgtgccaccggaagtc 900
DB 841 GTGTGTGCGTGTGATTTGATGAGAACAAATGAAATCCAAATGATGGCCACCCGAAGTC 900

QY 901 gacccggaaatccggtggaactcttcgtcttaagctacgaagtcgtgttcaagcctta 960
DB 901 GACCCGGAATCCGGTGAACCTTCTTAAACCTACGAGCGTGTTCAAAGCCTTAACCTA 960

QY 961 aaatctccgattctcaccgagcgaactaaatcaacccgagctcgagatcgaacttgat 1020
DB 961 AAATCTCCGATTCTCACCGGACGGAACCTAAATACCGGACGAGCTGAGATTAGCTTGAT 1020

QY 1021 cagccaacgattgatacagatttcgcatatcagagaacattcgtcgttcaacttgaccag 1080
DB 1021 CAGCCAACGATGATGACAGATTTCGCGATTACAGAGAACTTGTCGTGCTACCTGACCGAG 1080

QY 1081 caagtcgcttccaagctgcgcggaagatgataccgcggttgggtctccggttgaacaa 1140
DB 1081 CAAGTCGTTTCAAGCTGCGGAGATGATCCGCGGTGGCTTCGCGTGGTTTACGACAAAG 1140

QY 1141 aacaagtcgcaagatttcggaattttagacaataacgccaagattcactcgacaataag 1200
DB 1141 AACAAAGTGCAGAAATTCGGGATTTTAAACAAATACCGGAAGATTTCACGAACATTAAAG 1200

QY 1201 tggattgatgtccagatgtcttcgttccacttccttggaacgcttgggaagacgaaga 1260
DB 1201 TGGATTGATGTCTCAGATGTCTTGTCTTCCATCTCTGGAAGCTTGGGAAGACGACAGA 1260

QY 1261 acaagatgaagtcgtgtgaatgaaggtctgtatgaactcacaacgactcaatttcaagag 1320
DB 1261 ACAGATGAAGTGTGTGATGAGGTCTGTATGACTCCACCGAGCTCAATTTTCAACGAG 1320

QY 1321 tctgaacgaatcctcaagatgtctcgttctgaaatccgcgttaactcctcaaacggttga 1380
DB 1321 TCTGACGAATCTCAAGGTGTCTGTGTGAATCCGCTCAATCTCAATCTCAAAACCGGTGA 1380

QY 1381 tcaactgcgcgtccgatactctccaacgaagatcaacaagttcaactcgaagcaggatg 1440
DB 1381 TCAACTGCGCCTCCGATCATCTCCAAACGAAGATCAACAAGTCAACCTCGAACGAGGATG 1440

QY 1441 gttcaacagaacatgctgcgcgttaaaccaaatcgcgttacttggcttgaacgagccg 1500
DB 1441 GTTCAACAGAAATGCTCGCGCTTAACCAANATTCGCTTACTTGCTTTAGCCGACCG 1500

QY 1501 tggcctaaagctcagagatttcgctaaagtgtatcctaactcgtgagaagttaagaacat 1560
DB 1501 TGGCCTAAAGCTCAGAGATTTCGCTAAAGTGTATCTACACTACGAGAGATTAAAGAACT 1560

QY 1561 ctttaacgcgataaacggttaacgagagagagcctcgttcttcccggaagagagagag 1620
DB 1561 CTTTAAACGCGATTAACCGTTTACGAGAGAGCCTCTGTTTCTCCCGGAAGAGAGAGAG 1620

QY 1621 gaaagcgaagatatactcctcgttgcgttcaacggaagaagaacatgaaatcgaagtta 1680
DB 1621 GAAAGCAGAAAGTATATCTCTGTTTCTGTTTCAACGAGAGAAACATGAAATCGGAGTTA 1680

QY 1681 cagatagttaaacgcgtttagcttgaagttgaagaaaggtttaaacttcgctcaaggtt 1740
DB 1681 CAGATAGTTAAACGCGTTAGCTTGAAGGTTGAAGCAACGGTTAAACTTCCTCCAGGGGTT 1740

QY 1741 ccgtacggaattcaacggtatcatctcggagcggatgtatgtgggaagcaggtcgtgtga 1800
DB 1741 CCGTACGGATTTCACGGTATCATTCGAGCCGATGTGATTTGGCCAAACGAGCTCTGTGA 1800

RESULT 2
AB026549 3869 bp DNA linear PLN 24-OCT-2001
LOCUS

DEFINITION	Arabidopsis thaliana gene for neoxanthin cleavage enzyme, complete cds.				
ACCESSION	AB026549				
VERSION	AB026549.1 GI:16416373				
KEYWORDS	neoxanthin cleavage enzyme.				
SOURCE	Arabidopsis thaliana (strain:Col.) DNA.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucotsids II; Brassicales; Brassicaceae; Arabidopsi.				
AUTHORS	1 (bases 1 to 3869)				
TITLE	Iuchi,S., Kobayashi,M. and Shinozaki,K.				
JOURNAL	Characterization of neoxanthin cleavage enzyme from Arabidopsis thaliana				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 3869)				
TITLE	Iuchi,S. and Shinozaki,K.				
JOURNAL	Direct Submission				
FEATURES	Submitted (22-APR-1999) Satoshi Iuchi, RIKEN, Plant Mol. Bio.; 3-1-1 Koyadai, Tsukuba 305-0074, Japan (E-mail:luchi@rct.riken.go.jp; Tel:81-298-36-4359)				
source	Location/Qualifiers				
CDs	1..3869				
	/organism="Arabidopsis thaliana"				
	/strain="Col."				
	/db_xref="taxon:3702"				
	1831..3630				
	/codon_start=1				
	/product="neoxanthin cleavage enzyme"				
	/protein_id="BAB70609.1"				
	/db_xref="GI:16416374"				
	/translation="MASEFTAAVSGRMIGGNHNPPLSSOSSDLGYCSLPMASRVYTRKINSALHTPPALHPPOSSMSPAIVYPRKSESTKOMTLPORAAALDAAGSFVYHEKHLPIPLKTAIDPSVOIAGNFAPVNEOPVRNLPVYKLPDSIKGYIRNGANMLHEVYHIFHFDGGMVAVKFEHGSASYARFLTYNFEVDERLOGRVPEKAIAGELHGHTEIARLMFEFARAAGIYDPAPAGTGVAANGVLVFNGLRAMEBDLPYVOQTTPNGDILKTVRDEDFDGLDESLTMIAPKVPDSESGELFASLYVSGPVYKRYFSEDTKSPDVEIOLDPOTMMDHDAIETENFVYVDDQVFLPIMIKGSPVYDKMKVAFGLDKRYAESDSNKWIDARPCFPHFLMANEPEDETVVYVIGSCMPDPSIPNEDSNKLSVLSAERLNKTGESTRRPPIISNEDQVNLKAGNWRMISLRKTEFATLALAPPKVSGRAKYDITTESEVKHHLIGDMRNGGSEPLFPLRGESEDEEGVILICFVHDEKTIWSELTQVNAVSLEAEVAVKLPESRVYPGHGTGADDLAKOV"				
BASE COUNT	1155 a 845 c 773 g 1096 t				
ORIGIN					
Query Match	100.0%; Score 1800; DB 8; Length 3869;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 1800; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
0Y	1 atgagcttcttcaagcaagcgctcgcggttcttgaggagatgctctgtgcaatcatact 60				
Db	1831 atgagcttcttcaagcaagcgctcgcggttcttgaggagatgctctgtgcaatcatact 1890				
0Y	61 cagcgccgcatatcgctcttctcaaaagctccgcgacttgagattatgtactccttacctatg 120				
Db	1891 cagcgccgcatatcgctcttctcaaaagctccgcgacttgagattatgtactccttacctatg 1950				
0Y	121 gccagctgtgtacacgctaaagctcaatgtttcatctgtgcgttccacaactccctcagctctt 180				
Db	1951 gccagctgtgtacacgctaaagctcaatgtttcatctgtgcgttccacaactccctcagctctt 2010				
0Y	181 catttcctcctaagcaatcacaactcctccgcgcatctgtgttaagcccaaaagccaaga 240				
Db	2011 catttcctcctaagcaatcacaactcctccgcgcatctgtgttaagcccaaaagccaaga 2070				
0Y	241 tccaaactactaaacagatgaatttgttccaagaagcgggcgcgagcggtttgagacgcg 300				
Db	2071 tccaaactactaaacagatgaatttgttccaagaagcgggcgcgagcggtttgagacgcg 2130				
0Y	301 gagggtttccctgttcaagcagagaagatlaacccgcgtcttccaaagcgctgatactag 360				
Db	2131 gagggtttccctgttcaagcagagaagatlaacccgcgttccaaagcgctgatactag 2190				

QY	361	gtcaagatcgccggaatatctgtctccgtgaaatgaacagcccgctccgcgtaatctctccg	420
Db	2191	gttcgagatccgcgggaattttgtcccggtgaatgaacagcccgctccgcgtaattctccg	2250
QY	421	gtgtgtcggaanaactcccgatctcatcaagaagtgatagtgtggaacggaacttaacca	480
Db	2251	gtgtgtcggaanaactcccgatctcatcaagaagtgatagtgtggaacggaacttaacca	2310
QY	481	ctcaagcccggtgacaaagttcaaccattcttcgacgaagacgtatagttcaagccgtc	540
Db	2311	cttcacgagccggtgacaaagttcaaccattcttcgacgaagacgtatagttcaagccgtc	2370
QY	541	aaattcgaaacggtltcaagtagctacgtcttcggttacttaagactaaacggtttgtt	600
Db	2371	aaattcgaaacggtltcaagtagctacgtcttcggttacttaagactaaacggtttgtt	2430
QY	601	cagaagactcaatttgggttcgaacggttttcccaagaacatcgtgtgaagttcaagccac	660
Db	2431	caggaacgtcaatttgggttcgaacggttttcccaagaacatcgtgtgaagttcaagccac	2490
QY	661	accggtattgccgcgacatcatgtattctacgcgcaagctctgcagccgtlatagttcgaccg	720
Db	2491	accggtattgccgcgacatcatgtattctacgcgcaagctctgcagccgtlatagttcgaccg	2550
QY	721	gtaacaggaacccggttgtagtagtaagccggtttgtctatttaahgcccgttatgtgt	780
Db	2551	gtaacaggaacccggttgtagtagtaagccggtttgtctatttaahgcccgttatgtgt	2610
QY	781	atgtcggagagatgtattcaacttcaagatctcaatcaatcccaatbgaagatttaaaac	840
Db	2611	atgtcggagagatgtattcaacttcaagatctcaatcaatcccaatbgaagatttaaaac	2670
QY	841	gttggtcggttcgaattcttgatbgaacaaatagaatccacaatgtatgtgccaccggaagtc	900
Db	2671	gttggtcggttcgaattcttgatbgaacaaatagaatccacaatgtatgtgccaccggaagtc	2730
QY	901	gaccggaaatccggttgaaactcttcgcttgaagtagcaagctgttccaagacttaacta	960
Db	2731	gaccggaaatccggttgaaactcttcgcttgaagtagcaagctgttccaagacttaacta	2790
QY	961	aaatacttcgcatcttcaaccggaacgaactaaatcaaccggaacgttcgaattcaagcttat	1020
Db	2791	aaatacttcgcatcttcaaccggaacgaactaaatcaaccggaacgttcgaattcaagcttat	2850
QY	1021	cagccaagatgatagcacggttttcgattacgaagaactcgttgtctacttacaag	1080
Db	2851	cagccaagatgatagcacggttttcgattacgaagaactcgttgtctacttacaag	2910
QY	1081	caagtcgcttccaagctcgcgcggaatgatccgcggtgtctccggttgttcttacgacaag	1140
Db	2911	caagtcgcttccaagctcgcgcggaatgatccgcggtgtctccggttgttcttacgacaag	2970
QY	1141	aacaaagtcgcaagaattccgggaatttagacaabaatcgcgcgaagaattcatcgaaacttaag	1200
Db	2971	aacaaagtcgcaagaattccgggaatttagacaabaatcgcgcgaagaattcatcgaaacttaag	3030
QY	1201	tgtgatttgatccagaatttcttcgtcttcacatctctbgaacgcttgggaagaacgaaga	1260
Db	3031	tgtgatttgatccagaatttcttcgtcttcacatctctbgaacgcttgggaagaacgaaga	3090
QY	1261	acagaatgaagtcgtcgtgataaggttccgtatgaactccaccagaactcaatttccaagag	1320
Db	3091	acagaatgaagtcgtcgtgataaggttccgtatgaactccaccagaactcaatttccaagag	3150
QY	1321	ctctaacggaatctcaagaagttctcgtctctggaatccgccttaacttcaaaacggttga	1380
Db	3151	ctctaacggaatctcaagaagttctcgtctctggaatccgccttaacttcaaaacggttga	3210
QY	1381	tcaactcgcgcttcgacatcatctccaaagaagatccaaacgaactcaaccctcgaagcagagtg	1440
Db	3211	tcaactcgcgcttcgacatcatctccaaagaagatccaaacgaactcaaccctcgaagcagagtg	3270

Qy	1441	gtcaacagaacatgctcgccgtaaaccaaatcgcgttacttggtccttagccgagccg	1500
Db	3271	gtcaacagaacatgctcgccgtaaaccaaatcgcgttacttggtccttagccgagccg	3330
Qy	1501	tgccctaaagtcctcaggaattcgctaaagtcgactcactactgsgaagcttaagaacat	1560
Db	3331	tggcctaaagtcctcaggaattcgctaaagtcgactcactactgsgaagcttaagaacat	3390
Qy	1561	ctttacggcgatcgaacggttagcgaagagagcctcgtttctcccccgggaaggaagag	1620
Db	3391	ctttacggcgatcgaacggttagcgaagagagcctcgtttctcccccgggaaggaagag	3450
Qy	1621	gaagcgaagagatacatcctctgcttcgcttcacgacgaagacatggaataatcgaagta	1680
Db	3451	gaagcgaagagatacatcctctgcttcgcttcacgacgaagacatggaataatcgaagta	3510
Qy	1681	cagatagttaacgcgcgttagcttagaggttgaagcaacggttaaacctccgctcaagggt	1740
Db	3511	cagatagttaacgcgcgttagcttagaggttgaagcaacggttaaacctccgctcaagggt	3570
Qy	1741	ccgtacggaatttcacggtacattcctcggaagcgaatgattggcgaagcaagctgtgta	1800
Db	3571	ccgtacggaatttcacggtacattcctcggaagcgaatgattggcgaagcaagctgtgta	3630
RESULT	3		
AB028617/c		52232 bp	DNA
LOCUS	AB028617		linear
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, pl clone:MOA2.		
ACCESSION	AB028617		BA000014
VERSION	AB028617.1		GI:5041970
KEYWORDS			
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone_11b:Mitsui P1 clone:MOA2.		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.		
REFERENCE	1 (sites)		
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S.		
TITLE	Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty P1 and YAC clones		
JOURNAL	DNA Res. 7 (2), 131-135 (2000)		
MEDLINE	2 (bases 1 to 52232)		
REFERENCE	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E. and Tabata,S. Direct Submission		
AUTHORS	Submitted (09-JUN-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)		
JOURNAL	Address for correspondence: kaos@kazusa.or.jp		
	For the latest information on annotation of this clone, please see		
COMMENT	http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MOA2		
	Genes with similarity to proteins in the databases are described in		
	'product' or 'note' qualifiers. Genes that have no significant		
	protein similarity are described as 'unknown protein'.		
	The software programs used to predict genes include: Grail		
	(Informatics Group, Oak Ridge National Laboratory,		
	http://compbio.ornl.gov/Grail-1.3/),		
	GENSCAN (Chris Burge, MIT, http://CCR-081.mt.edu/GENSCAN.html),		
	NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of		
	Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and		
	SplicePredictor (Volker Brendel, Stanford University,		
	http://gremli.ni.zool.jastate.edu/cgi-bin/sp.cgi).		
	Genes encoding tRNAs are predicted by tRNAscan-SE		
	(Sean Eddy, Washington University School of Medicine, St. Louis,		
	http://genome.wustl.edu/eddy/tRNAscan-SE/).		
	This sequence may not be the entire insert of this clone. It may be		
	shorter because we remove overlaps between neighboring submissions.		
FEATURES	The 5' clone is MN21 and the 3' clone is MIE1.		
	Location/Qualifiers		

	source
1..52232	/organism="Arabidopsis thaliana"
	/strain="Columbia"
/db_xref="taxon:3702"	
/chromosome="3"	
/clone="MOA2"	
/clone_11b="Mitsui P1"	
join(1996..2132,2225..2349,2455..2501,2648..2725,2807..2931,3021..3087,3216..3278,3370..3439,3529..3629,3718..3822,3906..4091)	
/note="gene_id:MOA2.2"	
/codon_start=1	
/evidence=not_experimental	
/product="glycolate oxidase"	
/protein_id="BAB01334.1"	
/db_xref="GI:11994212"	
/translation="MEITNTEYDAIAIKKQLPKKVYVYVSGAEDQNTLDBNNRNRILFRRLIDVSKIDMTTIVLGFKISMPINVAIPAMQMAHPDEYATARAASAAGTIMITLSMATSVSEEVASTGPGIRFQLYVKNRVNEDQLVRAERAGFKALITVDPRLGRESIDIKNRFETLPNLTLKNFEGLDGKMKDEANDSGLSVYAGIDRLSMKDVOMLORTKLPILVKGVLTEGDARLAIQAGAGIYENHGAROLDVYPATISLSEVVKATOGRTVPYLDGVRKRGTDVFRKALAGASGIFIGRPVYVSLAEGAGYRKVLQMLRDEFELTNALSGCSRLKETSIRNHITTEMDTFRPSRL"	
complement(join(4591..4659,4763..4924))	
/note="gene_id:MOA2.3"	
unknown protein"	
/codon_start=1	
/evidence=not_experimental	
/protein_id="BAB01335.1"	
/db_xref="GI:11994213"	
/translation="MIGRRAGTNRVGRKRDSSLTRFVDSVYFFRLAEFEILFVLFMLITVIFPKDLTSREYNRILVERKEPQSGSDIMPF"	
complement(11870..13669)	
/note="gene_id:MOA2.4"	
/codon_start=1	
/evidence=not_experimental	
/product="9-cis-epoxycarotenoid dioxygenase"	
/protein_id="BAB01336.1"	
/db_xref="GI:11994214"	
/translation="MASFTATAAVSGRMIGSNHNPPLSSOSSDLSYCSLPMASRVTRKIVSSALHTPPALHFRKQSSNPRIYVYKPKKESNTQOMNIFQRAAAALDAAGFLVSEKHLHPKTPADPSVOAGNFAPVNEQVPVRNLPPVQKLPDSIKGVYVRNGANPLHEPVTHHFEFDGGVAVAFEHGSGASYACRFQTNRFVQERQLGPRVPKAGIELHGTCIARIMLFYARAAGIADVPFAHGTAVANAGLVYENGRLAMSEEDLPYQITPBGDLKTVGFEDPGOLESTIMAHKPVDPSEGEFLASTYVVSXPKLYKRFSPDGKSPDVEIQDPTMWHDEPITENFVYVDPDOOVYKRLPMIGSGSPYVYDKKVAIRFGIDKYAESNSNKKMIDAPRCFPHLMANNEEPEFDEYVYVYSGCMPPQPSIFPESNPNSKYSIEIRLNKTGSESTRPITISNEQOVNLEAGNKNRMRLKRTKRFATLALAEPMKVSQFASLVEEATVKKLPESRVPYGFHGFICADDLAQV"	
join(29980..30349,30421..30528,30616..30660,30841..30913,31021..31133,31239..31298,31376..31492,31588..31649,31733..31793)	
/note="contains similarity to RNA-binding protein"	
gene_id:MOA2.5"	
/codon_start=1	
/evidence=not_experimental	
/protein_id="BAB01337.1"	
/db_xref="GI:11994215"	
/translation="MATVTEPMDYVYVVDYKVEVSTSEKTIIDEGIKSSITTSKTEIESRDMHRLVAMEKKNLPKLEFPSSYDPRKKNQVAKANQFLPADDFETTKQSGEEFDLDQAKDDNTRKRRNYSQGRRLRIGRISAKORDSIRRVYVSDIDQSVTEBGLAOLFSCNGQVYDCRIGDPSHSLVFAVEFADQAGALSLGTMLGDFPVAVLPSKTAI LPVAPFLPSESDREMKRTITCTNIDKRVSDVNPESACGEVTRIRLLGDQHLSTRIAVEFLPADSALSNLNSGMYVSSQPIRVSPTVPRRTTRPPSN"	
complement(32182..36456)	
/note="gene_id:MOA2.6"	
/codon_start=1	
/evidence=not_experimental	
/product="disease resistance complex protein"	
/protein_id="BAB01338.1"	
/db_xref="GI:11994216"	
/translation="MANSYLLSCANVYERINTSQELVELCKGSSALLKRLKVALV	

TANPLADADORAENHREVKWHLTGKIDAEFOEDILDEQTEALRRRVAAAGLGG
LPONMAGREAIQKLEPKMEKVVRLLEHHVKNHIEVIGLKEYSETREPMOQARSP
DDLPGRLVGRVDEKDLALVNLILSDDEISTGKPAVIVSGMPGKKTIEIVANDR
VTHREPKMISAGINNVTVYKAVIQDITSSAVNTEDPSPIDQIQKTLSCRFLI
LWDFWSESDSEWSEFOVATDAEGSKIVUTKSELVSTVAKEKTIYOMKMLNEEC
WELISRAFGNISVGSINOELGIGKIAQCKGLPLAARAIAHSLKSPKMDWYAV
SKFSSYTNLILPVLYKLSYDLPOLKRCFALCSIFPKGAFDEELVLMALMADLY
QPPSSRLNEDIGNVLDLVAOSFPORLDTMTSFMKMDLMDLAKAVSGFCRLD
DNIPETPTTRHFSRSRSCDASVAFRISQAEFLRTILPENSPISLISQIPEKYN
PLNALSGLRILSLSHQITNTPKSLKGLKILRLDSSRIKILPEFVCTICQIQL
LLSNCRDLTSLPKSIABLINDRLDLVGTPEVMPPEIKLRSQKLSNVIQGLSQA
GLHLEKLSHLRGLTIRLSELQNAFASEAADAGAKRPFLLDGLILKKTAVGSGPVS
FNAIACQKQEVRLMELEPHPKLTFECISVGGAGSPKSGSEFGITSVLSQNLICI
SLPEVQGLPSLKYLSIEKFNILQVGLDFEFGENNSRGVPEOSIOLIKFYGMFMDM
ICEPELDEIPLOKILITORCPLRKKPEPGLPSTETVTISDCLRAVSGENSFERS
LTNIPESPATIPSMRRELSPTCNPKSDASTSQOPFASSOSQNDNEVYTSLSLS
LPADQJEDDQJETQIGSLPQOPEEPVISAARSIGTISDIPSLYSMTSISLVPPE
KNEGSLIPGSSSYQHQYGIKISVSPRSSEAIKPSQYDDEDMEXLKAYDLSHME
LPONLSLHSDSCGLSLPENLIESYPNHELILICHSELESFSGHPPTLITLYI
RDCKLNFTSLQOTRYSOLEYLFIGSSSNLVPFLKPSLSIRDCSEFTF
SIHAGIDDRILALESLERDCPNLEPPOGLPDKLSSMLNSCKLOALPEKLFGL
TSLISLTIKCPLEETIPGGGFPSNLTICISLQDKITPRLEMLRDLNRLNLEIDG
GNEDIESFPEGGLPKSVSDRISRENLKTLNKGKPHDKALTEIMELSGCDKLQISI
DEDPLPLSCLRISSCSLTETFAVEVEFEKVLNIPVEIDELIFS"
38132. .41296

CDS

/note="gene_id:MOA2.7"
/codon_start=1
/evidence=not_experimental
/product="disease resistance complex protein"
/protein_id="BAB01339.1"
/db_xref="GI:11994217"
/translation="MTGIGEMFLAFLQALFOVLVSEPFNSFEKREINLELRLST
ALLITPAVLIDAEKQITNPVEKWKWMLLDVYVHADADILDALITELRLNIGNESS
SNRLROGRMSIGDPLDGSSEHLEFLEKVIIRLELAQSRLNLIKELTAMLPKOR
LPITSLVDESEVGRDDKDEIMRFLPEMGKDITIVAVLGVGGKTLISQTLN
DOHVRSTGKVAHVSEEDVEKIRKVESVTSRCEFDLDVLQIKLERLQIG
LPFLVLDDLMENFADMDLROPIHAAQGSQILVITRSRQVRSIMCAVAVHMLQPL
SDGCMGLFMKTYVGNQPCILREIGLARIVKCGGLPLAVTLGGVLEFEGKYLE
MERVLSRIMDLPADKSNLLPVLEVSYYIPAHKRCFVCSITPKHGAPEKDEVL
MMAEGTIOQRSSKLELIGNEVSELESLSLOKTRVYMHQFIMELIASQAFES
SKPEDCKLOVSEKTRILSLRQYAEPMERALEKFLTFLPLSTLNSRSCCDD
QNVSEKLPULTRLVLSLHYKLAIRLPPEFNI SHARFLDSITELKPKSLCTM
YNLOTLILSYCSLKEPLTDISNLINRYLDLIGTKLRMPRRGRKLSQTLTFEY
SASDGRSISELGGIHDHGLKILVELOVVADVAANLNSKHLEIDVMTGSS
SEENNTNPHRTONAEEVEKLRPHRIEKLAEYKSRPDMISDSFSITVIR
ECQCTSLPSIGOLPCLKEHLSGMGLSIGKRGVGLRQDPODPESTELTIR
DNLPDMQEMIDVYTRGDLRPSLKLPILEKPELTGTPPTLPBLSLHITKGLDLE
QPDHETSYRLQTLSTKSSCDTLVKFPLNHPAMDLEVDQCTSLVLESNHLRG
PNALRLRINDCOMQLPKLMLPOMLOYITNCRYLROPMEOQPOYHHPQHLPRS
NVSGSPKSHGSHRSYSSSRSD"
42345. .42872
/note="gene_id:MOA2.8
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01340.1"
/db_xref="GI:11994218"
/translation="MSDLVATVAVAFVAVLTVFTLLSCIMGKDYHSPPLRPGOLR
MPYVTRKDXDGGGRTAISVAGTDIYGSYGSGGHHGRGRTITSDTKLPSPPPSP
PRVQRPKSVSRKIDRELIDGTAMLVASSLITGSSASSCSGGSGSHHGCGGGGGG
GGLGGGGGGGGCGG"
complement(join(42980. .43465,43561. .43668,43952. .44119,
44215. .44460,44539. .44677,44759. .44977,45060. .45754))
/note="gene_id:MOA2.9"
/codon_start=1
/evidence=not_experimental
/product="terpene synthase-like protein"
/protein_id="BAB01341.1"
/db_xref="GI:11994219"
/translation="MKPKVRMLKSSKNDSMERIRLTHLMLNLTAVAFETEDIEIG
QARGNDDITAKENDLETISTMEVYRRLRGTYMCCSISGLDSHICHTIYNLADFRT
VDAPNRKGEDGRKESLAEIDRGMLQDYEAAHGTSESDIDALSTTRRLSLVS
NHATASPHLSKHIONALYRKYRHNLEILVAREYISFEQEDDEHDETLKPAKLNFY

CDS

CDS

COLHYIOELNDLTMMKRELDLASKLPYIRDRIVEVFGALATYEPREYSIGRIIVTKI
TMATVENDNCDADGLPEVTSILVDSORMDLGIKLPKSVKLVFGEVETLEIEIO
EMRPOGRSRIYOVAVDELKIKGLKALISKMARSHAPFEVEFERMOSHFAAY
SFIAMEBCDENQTEEMTKSRKMEBALNVRIRKINDIPEQEMSREVAKGLNCHYK
QHGSKEEAIGENMKITSNYKILIMEELVTTAPRPLLEFMDLNSKPLITKTLTM
FRVNDVGTVEVHLKLTETESRGNVQDELYMKQHVTEKESAOELRKNKMDYKXV
MEEMNTNHDHLPQGVFLRCHNIARIIFPVYFVGVGYCDPKKINEMTSLYIHPINP
FIMWNTNHOXSRSIRGCCIEIKKKKVVLSLLIHRYSNVLVHRCFVFLYNSIFLC

Query Match 100.0%: Score 1800; DB 8; Length 52232;
Best Local Similarity 100.0%: Pred. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	1	atgagcttttcacggaacgctgagcttcttcggagatgagcttggcgaatcaact	60
Db	13669	ATGCGTCTTTCACGCGCAACGGCTCGGTTCTGGGAATGGCTTGGGCAATATACT	13610
Oy	61	cagccgcacatatacgtctctcacaagctcgcagcttgaattttagctccttaccatg	120
Db	13609	CAGCGGCATTAATCGTCTTTCACGCGCAACGGCTCGGTTCTGGGAATGGCTTGGGCAATATACT	13550
Oy	121	gccagtcgttcacacgtaagctcaatgtttcattcgtcgtcctcaactcctcagctctt	180
Db	13549	GCCAGTCGTTCACACGTAAGCTCAATGTTTCAATGGCGCTTCAACATCTCCACCTGTT	13490
Oy	181	cattccctcaagcaatcaatcaactctccgcacattgttgaagcccaagaacaaqaa	240
Db	13489	CATTTCCTTAACATATCATCAACACTCTCCGCAATGTTGTTAAAGCCCAAGCAAGAA	13430
Oy	241	tcacaactaacaagatgaattgttccagaagagcgcgcgcagcgttggagcgcgcg	300
Db	13429	TCCACACTAAACAGATGAATGTTTCCAGAGAGGCGGCGGCGGCGTGGACCGGGG	13370
Oy	301	gagggttccctgttaagcacaagaaagctacacccgcttctcctaaagcgttgaacctg	360
Db	13369	GAGGGTTCCTGTCAGCAGCAGAGAGCTACACCCGCTTCAAAACGCGCTATCTGAT	13310
Oy	361	gtccagtcgcgcgaagaaatttgcctcgcgttgaatgaacagccgctcgcgcgtatcttcg	420
Db	13309	GTTCAAGATCGCGGAAATTTGCTCGGATGAATGAACAGCCGCTCGGCGATATCTTCG	13250
Oy	421	gtgtcgcgaaaaactccgcattccatcaacaagagtgatgtgtgcgaacgaggaacc	480
Db	13249	GTCGTCGGAACAACTCCGATTCATCAAAAGAGATGATGTCGCAACGAGACTTAACCA	13190
Oy	481	cttcacgcgcgcgttacagcgttaccacattcttcgcgcgcgcgcgcgcgcgcgcgc	540
Db	13189	CTTCACGAGCGGCTACAGGTCACCACTTCTTGCAGGAGAGAGGATGATGTTACCGCGTC	13130
Oy	541	aaatcgcgaacacgttcacgtacgtacgccttcgcgcgttcaactcaaacacgcgtttgt	600
Db	13129	AAATTCGAAACGAGTTCACTACGTACGCTTCGCGGTTTACTACACACTAACCGGTTTGT	13070
Oy	601	caggaacgtlcaattggttcgacgcgcttccccaagaacgcgtgtagcttcaacgcca	660
Db	13069	CAGGAACGTCAATTCGTCGACCGGTTTCCCAAGCAATGAGGCTTACCGGCGCAC	13010
Oy	661	accggtatgcgcgcgcctatgtatcttaacgcgaagctgcagccgcgcgcgcgcgcgc	720
Db	13009	ACCGGTATTCGCCGACTCATGCTATTCACGCAAGCTGACGCGGATATTCACACCG	12950
Oy	721	gcacgcgaacgcgtgtagcgaacgcgcgttgcgtcattcaacgcgcgcgcgcgcgcgc	780
Db	12949	GCACGCGAAGCGGTGCTACCTAACCGCGGTTTGGCTATTTCAATGCGCGGTTATGGCT	12890
Oy	781	atgtcgaagatgattacacttcaacgaatcagatcactcccaatggaatttaaacac	840
Db	12889	ATGTGGAAGATGATTACCTTACCAAGTTCAGATCCTCCCAATGGAATTAATAAAC	12830
Oy	841	gtgtgctgcttgcattttagatgacaatttagaataccaatgattgcccacccgcgaagtc	900
Db	12829	GTTGTGCTGCTGATTGATGAGACAAATTAATGACCAATGATGATGATGATGATGATG	12770

QY 901 gaccggaaatccggtgaactcttcgtttaagtaagtaagtggtttcaagcttacccta 960
 |||||
 Db 12769 GACCCGGAAATCCGTTAACTCTTCGTTTAAAGTACGAGCTGCTTCAAGCCCTTACCTA 12710
 QY 961 aatacttcgattctcacgcgacggaactaaatccacgagctcgagattcagcttgat 1020
 |||||
 Db 12709 AATATCTTCGATTCTCACCAGGAGAACTAAATCAACGAGCTTCAGATTCAAGCTTGGAT 12650
 QY 1021 ccgcaacgattgatacgaatttcgcatattacagagaacttcgtcgtctgactgacag 1080
 |||||
 Db 12649 CAGCCAAACGATGATGACGATTTCCGGATTACAGAGAACTTCGTCGCTGACTGACACAG 12590
 QY 1081 caagtgttttcaagcttcgagagatgatcccggtgtgtctcgtgtgtttacgacaag 1140
 |||||
 Db 12589 CAAGTCGTTTTTCAAGCTGCGCGGAGATATCCGGGTGCTTCGCTGTTTAAACGCAAG 12530
 QY 1141 aacaagtcgcaagatttcggtatttagacaataacacgcaagattcatcgacaataag 1200
 |||||
 Db 12529 AACCAAGTGCAGAGATTCCGGGATTTTAAACAATACGCCGAGATTCAATCGAATTTAAG 12470
 QY 1201 tggattgattcagattcgtcttcgtcttcacatctctggaacgcttgaggagacagaa 1260
 |||||
 Db 12469 TGGATTGATGCTCCAGATTGCTTCTGCTTCCATCTCTGGAAGCTTGGGAAGACAGAA 12410
 QY 1261 acagatgaagtcgtctgtagaggtctcgttagatccacacgaactcaatttcaacgag 1320
 |||||
 Db 12409 ACAGATGAAGTCTGCTGTGATAGGGCTCTGTATGACTCCACCACTCAATTTTCAACGAG 12350
 QY 1321 tctgaagaagaatcgaaggtgtcctgtctgtaaatccgcctgaatctcaaaacggttga 1380
 |||||
 Db 12349 TCTGACGAGATCTCAAGAGTGTCTGTCTGAAATCCGCTGAAATCTCAAAACCGGTGAA 12290
 QY 1381 tcaactgcgcgttcgatactctccacagaagatcaaacgaactcgaacgagagatg 1440
 |||||
 Db 12289 TCAACTGCGCTCCGATCTCATCTCCAAAGAAATCAACAAGTCAACTCGAAGAGGAGATG 12230
 QY 1441 gtcaagaagaacatgtctggtccggttaaaccaaatctgcttacttgcttagcgcgag 1500
 |||||
 Db 12229 GTCMAAGAAACATGCTCGCGCTAAACCAATTTGCTTACTTGGCTTAAAGCCGCGCG 12170
 QY 1501 tggccaaagtcgaagattcgtctaagtgtatctcactactcgagaagtaagaacat 1560
 |||||
 Db 12169 TGGCCCTAAAGTCTCAGAGATTCCGTTAAAGTTGATCTCAGTACAGAGAACTTAAAGAACT 12110
 QY 1561 cttaacgagcaataccgttaccgagagagagcctctgttctcccccggagagagagagag 1620
 |||||
 Db 12109 CTTTACGGGATTAACCGTTACGAGAGAGAGCCCTGCTTCTCCCGGAGAGAGAGAGAG 12050
 QY 1621 gaagacgaagagatacctctcgttctcgttcaacgagagagacatgaaatcgagatta 1680
 |||||
 Db 12049 GAAGAGGAAGGATACATCTCTGTTCTGTTCAAGAGAGAAATCGAGATTGA 11990
 QY 1681 cagatagttaaacgcgtttagcttagaggttgaagcaaggttaaatcttcgtcaaggtt 1740
 |||||
 Db 11989 CAGATAGTTAAACGCGCTTAGCTTAGAGGTTGAAGCAACGTTTAACTTCGTCACAGGCTT 11930
 QY 1741 ccgtacgagattcaacggttacatctcgttctcgttcaacgagagagatttgcgagagcaggttctgtga 1800
 |||||
 Db 11929 CGTACGAGATTCAACGAGTACATTCATCGAGCGAGATGATTGGCGAAGCAGGTCGTGTA 11870

RESULT 4
 AY056255 2331 bp mRNA linear PLN 30-SEP-2001
 LOCUS Arabidopsis thaliana putative 9-cis-epoxycarotenoid dioxygenase
 DEFINITION (MOA2.4/AT3g14440) mRNA, complete cds.
 ACCESSION AY056255
 VERSION AY056255.1 GI:15810432
 KEYWORDS FLI_CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
 AUTHORS
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 2331)
 Yamada,K., Banh,J., Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G.,
 Yu,S., Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
 Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
 Kim,C., Koeseema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
 Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
 Shinn,P., Southwick,A., Tracy,S.E., Shinozaki,K., Davis,R.W.,
 Ecker,J.R. and Theologis,A.
 Full length cDNA of gene MOA2.4/AT3g14440 (GI:11994214)
 Unpublished
 2 (bases 1 to 2331)
 Yamada,K., Banh,J., Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y.,
 Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H., Cheuk,R.,
 Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
 Kawai,J., Kim,C., Koeseema,E., Lam,B., Lin,J., Meyers,M.C.,
 Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
 Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
 Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
 Direct Submission
 Submitted (12-SEP-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
 Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S.,
 Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S.,
 Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C.,
 Koeseema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tracy,S.E., Davis,R.W.,
 Ecker,J.R. and Theologis,A.

COMMENT
 JOURNAL
 The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Yamada,K., Banh,J.,
 Banno,F., Dale,J.M., Goldsmith,A.D., Lee,J.M., Onodera,C.S.,
 Quach,H.L., Tang,C.C., Toriumi,M., Yamamura,Y., Yu,G., Yu,S.,
 Bowser,L., Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C.,
 Koeseema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
 Palm,C.J., Shinn,P., Southwick,A., Tracy,S.E., Davis,R.W.,
 Ecker,J.R. and Theologis,A.

FEATURES
 source
 1..2331
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL08-11-H16 (R11094)"
 /note="This clone is in a modified pBluescript vector
 (FLC-1) as a BamHI/XhoI insert.
 ecotype: Columbia"
 1..2331
 /gene="MOA2.4/AT3g14440"
 1..122
 /gene="MOA2.4/AT3g14440"
 123..1922
 /gene="MOA2.4/AT3g14440"
 /codon_start=1
 /evidence=experimental
 /product="putative 9-cis-epoxycarotenoid dioxygenase"
 /protein_id="RAFL07104.1"
 /db_xref="GI:15810433"
 /translation="MASTPATAAAGRWLGNHTPOPLSSQSSDLGYCSLPMASRY
 TRKLNVSSALHTPAPLHPKQSSNSPALVYVCPKAKESNTKQNFQRAAARAAALAE
 FLVSEKRLPIKTPADSPVOJAGNPAVNEQPVRRNLPVYKELPDSIKGYVVRGANG
 LHEPYTGHHFFPDGGMNAAVAFEHCSASYACRFQTRNRFVBERDLGRVPFPAKAGELH
 GHTGARLMLTFYARAAGTVPDPAHGTGVAAGLVYFNGRLAMSDDLTPVQITPNS
 DLKTVGRDFDGOLESTIMAHKVPDSEGLFALSYDVVSKPYLKYRFPSPDGTSPD
 VEIQLDQPTMHADPAITENFVVVDDQVYFELPMLILGSPVYVDKNVAFGLDKS
 AEDSNKIIDAPDCEFHLMNMEDEPETDVAIVIGSGMPDPSIPNBSDNLSVLS
 EIRLNKKGESTRRPRTISNEQOYNLEAGVNRNMGCTKTKFATIALAEPMPKSGFA
 KYDLITGKVKHHLIGDNRNGSEPLFPLREGEDEBGLTLCVHDEKTKWSBLQIVNAV
 SLEVAIVKLPSRVYPYGFHFTIGADILAKOV"

```
misc_difference 1234
/gene="MOA2.4/AT3g14440"
/note="compared to genomic sequence resulting in an amino
acid sequence difference"
/replace="g"
1923. 2331
3'UTR
/gene="MOA2.4/AT3g14440"
misc_difference 2059
/gene="MOA2.4/AT3g14440"
/note="not present in genomic sequence"
misc_difference 2312
/gene="MOA2.4/AT3g14440"
/note="not present in genomic sequence"
misc_difference 2315
/gene="MOA2.4/AT3g14440"
/note="not present in genomic sequence"
BASE COUNT 615 a 560 c 533 g 623 t
ORIGIN
```

Query Match	99.98%	Score 1798.4	DB 8	Length 2331
Best Local Similarity	99.98%	Pred. No. 0		
Matches 1999, Conservative	0	Mismatches 1	Indels 0	Gaps 0

OY	1	atgctctcttcaaggaacgaagcgctcggtttctcgggagatgcttggtagcaatactact	60
Db	123	atggctctcttcaaggaacgaagcgctcggtttctcgggagatgcttggtagcaatactact	182
OY	61	cagcgccattatcgctcttccaagctccgaacttgatattgtagtccctactatg	120
Db	183	cagccgcattatcgctcttccaagctccgaacttgatattgtagtccctactatg	242
OY	121	gccagctcgctgacacgtaagctcaatgctttcatalctgagcttcacacatccctcagctct	180
Db	243	gccagctcgctgacacacgtaagctcaatgctttcatalctgagcttcacacatccctcagctct	302
OY	181	catttccctaaagcatcatcaataactccccgcattgtgtttaagccaaagccaagaa	240
Db	303	catttccctaaagcatcatcaataactccccgcattgtgtttaagccaaagccaagaa	362
OY	241	lccaacacttaaacagatgaatttctccagagacgagcgagcgatgtgacgagcg	300
Db	363	tcacaacacttaaacagatgaatttctccagagacgagcgagcgatgtgacgagcg	422
OY	301	gaaggcttcccttgctcagccacgagaagactacacccgcttccctaaacgagtgatcc	360
Db	423	gagggcttcccttgctcagccacgagaagactacacccgcttccctaaacgagtgatcc	482
OY	361	gttcagatccgcggaaatttgcctccggtgtaatgaagaacgcggtccgggttaacttc	420
Db	483	gttcagatccgcggaaatttgcctccggtgtaatgaagaacgcggtccgggttaacttc	542
OY	421	gtgtgcgaaaaacttcccgattccataaagaggtgatactgtgcgaacggagtaaccca	480
Db	543	gtgtgcgaaaaacttcccgattccataaagaggtgatactgtgcgaacggagtaaccca	602
OY	481	cttcaagcgccggtgacaggttcacacacttcttcaacgaaacggtatggttcaacgcgtc	540
Db	603	cttcaagcgccggtgacaggttcacacacttcttcaacgaaacggtatggttcaacgcgtc	662
OY	541	aaattcgaaacaggtltaagctagctacgcttgcggttttaactaagactaaacggtttgt	600
Db	663	aaattcgaaacaggtltaagctagctacgcttgcggttttaactaagactaaacggtttgt	722
OY	601	caggaagctcaattgggtgtgacacggttttccccaagcatcggtgagcttcaagcgac	660
Db	723	caggaagctcaattgggtgtgacacggttttccccaagcatcggtgagcttcaagcgac	782
OY	661	accggtatgcccgaactatgctatcttbaacggaagcttcaacggtatagtcgagccg	720
Db	783	accggtatgcccgaactatgctatcttbaacggaagcttcaacggtatagtcgagccg	842
OY	721	gaacaggaacacggtlgtagcttaaacgcggttggcttcaattcaatgacgcggtattagct	780

Db	843	GCACACGGAACGGCTGTACTTAACGGCGGCTTGGCTATTTCATATGCGCGGTATTGGCT	902
Qy	781	atctcggaagatgatatttaccttaaccagaattccagatcattcccaatggagaatttaaaacc	840
Db	903	ATGTGCGAGATGATTTTACTCTTAACCAAGTTCCAGATCTACTGCCATGAGAGTTTAAAAACC	962
Qy	841	gttgcctcgatcgatttgatgacaattagaatccacaatgatgtgccaccggaaagtc	900
Db	963	GTGTGCGGTTGCGATTTTATGATGAGCAATTAAGATCCACAATGATTTGCCACCAGAAATGC	1022
Qy	901	gaccgggaatccggfbaactcttcgctttaagctagaagctcgattccaagccttcccta	960
Db	1023	GACCCGGAAATCCGGTGAATCTTTCGTTTAAGCTACGACGCTCTTTCAAGCCCTTACTTA	1082
Qy	961	aaatactccgaattctcacgcgcgaacaaataacccgaagctcgagatctgaattcgat	1020
Db	1083	AAATATTTCCGATTTCTCACCGACGGAAGCTAAATATCACCGGACCTCGAGATTACGTTGAT	1142
Qy	1021	cagccaagaatgatgcagaattcgcatatacagagaactcgtcgctgcattgcacga	1080
Db	1143	CAGCCACACATGATGCACATTTCCGGATTACAGAAACCTCTGCTCGACCTTGACAG	1202
Qy	1081	caagtcgatttcaagctgcgcgagatgatccggggtggtgttcctcggtgtttcagcaag	1140
Db	1203	CAGTCTGTTTTCAAGTGTGCGCGGAGATGATCTCTGGGGCTCTCGGCTTTACGCACAG	1262
Qy	1141	aacaaagtccgaagatctcggaattcttagaacaatacgcgcgaagatcattcgaaattaa	1200
Db	1263	AACAAGTGTGCAAGATTCGGGATTTTAGACAAATACGCCAATATCTCGAATATTAAAC	1322
Qy	1201	tggaattgatgtccagaattgtcttcgtctccatctcttgaaagctctgggaagccaaga	1260
Db	1323	TGGAATTGATGCTCCACATTTGCTTCTGCTTCCATCTCTGGAAACCTTTGGGAAGCCAGAA	1382
Qy	1261	acgaatgaagatcgctgtgtataggtgctcttataagaccacacgaatccaattttcaacga	1320
Db	1383	ACAGATGAAGTGTGTGTGATAGGGTCTCTGTATACTCCACACACTCAATTTTCAACGAG	1442
Qy	1321	tcctgacgaagaatctcaagaagtgtccctgtctgaaatccgcctgtaatctcaaaaccggtga	1380
Db	1443	TCGTGACGAGATCTCAAGAGTGTCTGTGTAATCCGCGCTGAATCTCAAAACCGGTGA	1502
Qy	1381	tcaactcgcgcgttcgatcatcttccaaacgaagaatcaacaagtcaactctgaagcaaggatg	1440
Db	1503	TCAAATCGCGCTCCGATCTATCTCCAAAGAAATCAACATCAACTCGAAGACGGGATG	1562
Qy	1441	gtcaacagaagaacatgtctgcgcgttaaacccaatatctgcattacttgcttagccgcgcg	1500
Db	1563	GTCACAGAAACATGCTCTGGCGCTAAAACCAAAATTGGCTTACTTGGCTTTAGCCGAGCCG	1622
Qy	1501	tggtctaaagttctcagaattcgcgtlaaagtgtgactctactacttggaagaagttaagaacat	1560
Db	1623	TGGCTTAAAGTCTCAGAGATTTCGCTAAAGTTGATCTCACTACTGTGAGAGAGTTAAGAAACAT	1682
Qy	1561	ctttacgcgcgataaacccgttaacggaagaagacgtctgtttctcccgcgagaagaagaagag	1620
Db	1683	CTTTTACGGGAGATTAACGTTTACGAGGAGAGACCTCTGTTTCTCCCGAGAGAGAGGAGAG	1742
Qy	1621	gaagaagaagataactccctctgttccgttccacgaagagaagacatggaatctggaattta	1680
Db	1743	GAGAGAGGAAGATACATCTCTTCTTTCGTTCCAGACGACAGAAACATGGAATGGAATTTA	1802
Qy	1681	cgatagttttaaagccgtctgaagcttagaaggttgaaagcaacggtttbaacttccgtcaaggtc	1740
Db	1803	CAGATGTTTAAACCCGCTTGAAGCTTAGAGGTTGAAGCAACGTTTAACTTCGCTCAAGGTT	1862
Qy	1741	ccgtacgcgatttcaacggtatcatctacgcgaagccgatgatattggcgaagcaagtcgtgtga	1800
Db	1863	CCGTACGGATTTTACCGTATCATCTACGAGCCGAGATGATTTGGCGAAGCAGGTCTGTGA	1922

RESULT 5

AC013430		95769 bp	DNA	linear	PLN 26-MAY-2000
LOCUS	AC013430	Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome I, complete sequence.			
DEFINITION	AC013430				
ACCESSION	AC013430				
VERSION	AC013430.5	GI:8096768			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopses.				
AUTHORS	1 (bases 1 to 95769)				
REFERENCE	Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C., Shin,D., Altafi,H., Bel,Q., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzales,A., Hansen,N., Howng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Tortumt,M., Vaysberg,M., Yu,G., Federspiel,N.A., and Ecker,J.R.				
TITLE	Genomic sequence for Arabidopsis thaliana BAC F3f9 from chromosome I				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 95769)				
AUTHORS	Ecker,J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-NOV-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	3 (bases 1 to 95769)				
AUTHORS	Ecker,J.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-FEB-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA				
REFERENCE	4 (bases 1 to 95769)				
AUTHORS	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Tortumt,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,U.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-MAY-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA				
REFERENCE	5 (bases 1 to 95769)				
AUTHORS	Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bel,B., Chin,C., Chlou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N., Howng,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharasy,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A., Thaveri,A., Tortumt,M., Vaysberg,M., Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,U.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-MAY-2000) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA				
COMMENT	On May 26, 2000 this sequence version replaced gi:6921155.				
FEATURES	Location/Qualifiers				
SOURCE	1..95769				
	/organism="Arabidopsis thaliana"				
	/db_xref="taxon:3702"				
	/chromosome="1"				
	/clone="F3f9"				
	join(71..402,574..649,733..933)				
	/note="similar to oxidosqualene cyclase dbj BA33462.1"				
	/codon_start=1				
	/evidence=not experimental				
	/product="F3f9.1"				
	/protein_id="AAF171792.1"				
	/db_xref="GI:8052528"				

CD5

"/translation="MGVQLTLPCETYSPEPFSFLLEFLFNKAVFLILISVLTGTV
FLELKEVDEIYPVAKVEDGESSTHTRTMRARGEEFLDSRLCPRMVGLKESLDLSSSS
FLEWVSSSEFIELEIADAEHREILLYIRHLNDGMLGHAHEGSMFCFALNTYI
CLRIREGDEGREGACKLSTSMPIVFFLLIYNNSCEFSPI"
complement(join(3438..3746,3868..4032,4110..4664))
/note="similar to F-box domain gb|AA14684.1"
/codon_start=1
/evidence=not_experimental
/product="F3f9.2"
/protein_id="AAAF71813.1"
/db_xref="GI:8052549"
/translation="WLASMTOSLTSMKHTKLLILANENGRAPPGEDNVLKSTETPN
HPSLRLLSYOYLIDDAOFNNCKNLKLDLFOIJAEDVENEYISSCPSELYLIDLI
IFPNSGHLIKEDHKTLRILMSCSNOIDSEVRARLDLTSLEYIPCSNDVLEIPL
OGRWNVWAGRLPLPTSLINISCPPOKSNMVMKDYATSPASLSYDLKPREVL
EVLKRLIAWTEKRIEVEISFKNKSNLPGEDNNSGGKONNLAMKADPAPADRVDT
TMHNPFKSNKEQFALASRFVAKQVNMKMKIKTSFDEKKKKEIEAAVAKLKLPRKN
EELSTEDCF"
join(7138..7400,7491..7887)
/note="unknown protein: similar to Esp gb|A199747.1"
/codon_start=1
/evidence=not_experimental
/product="F3f9.3"
/protein_id="AAAF71793.1"
/db_xref="GI:8052529"
/translation="METSILIKLSLCSLVVGSYAQAPAPANNSNGFRPGCDHYE
CTYKLVENGYSFEIRMTDALIMISTSPISLSTMTQATKTFERLNTYIEGDKSNVK
NMVAAPVIAQATPGRSYVTVSLYLRKKNQPPADADLVHSTKPTVAARIGGVYS
NMVADDEAALMESLRSNMLPIEKSQKGLPAFLVAVNPSTITRVLINELVNPNN
M"
CD5

join(8524..8759,8846..9287)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.4"
/protein_id="AAAF71794.1"
/db_xref="GI:8052530"
/translation="MEALSIFFKSLFTLLTSLGSGDLISLPPNCRNAECPSEYVHA
NGYELHRTNWTWISTEPIODISLNEAGSGNMOLSDYNNKNDYHOREIAPLPTT
OYSONLSPIFVSFEFPAKAFOPDPNPGNLIHVORNSRVAAKQISGVADHKIGKQVA
EKLASLQGVAKALIEKSRFEGVGSAAATYVAQFSNPFQNSQVYNELWPFEMDEE
TVSIYVO"
join(13444..13814,13956..14310,14399..14662)
/note="similar to dihydrogenase protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.5"
/protein_id="AAAF71795.1"
/db_xref="GI:8052531"
/translation="MAVLSIKPVAIPKSGSLPIVIDMSDEPSKHALVAKADEGFEKVV
IHNGVSAELVSLHEHETVDFPSLSEKTYOYVAPFGNGVIGKRGNDVGWVEFLFN
ANDHQSQSPSLIKSPGTGRNALIEYTSVRKMTDGVLEKIRIDGIGLIPRNTLSL
VSDMDTSLIRLNHPPCPLSNKTKNGKAVNGICEHTDDPIISLVLSNNTSGQILN
NGSNIIVPBDITSEFFNVGDSLOYMTNGKRSVRRHRLAACKSRVSMYTFGSPSLT
QRIALPQLDINEDRLYEFTWSELYKSTYNSRLSDNRLOQFERKTIKLN"
complement(join(15755..16715,16810..16829))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.6"
/protein_id="AAAF71812.1"
/db_xref="GI:8052548"
/translation="MKPRGLIGRSELPOROSPRLTRSLSTSSDPHLISRPITDRSP
KGLDRSPRSRGPTDPLISOKIKLGRSGESLQIGAOEELIKTKOILAKAPAKR
AOEELHRRKSKSPNTPAPERDIPEDGHOETDVEVLDEKAKSEKTKNDLASKEDQ
INVLAKRLYDLERKSVLSSENETLKDQILKTDIVEMSCAKADELSKVSQIGEELE
ESNETTAKIKLESEVGEAKETLAEAKMKLVQVEGKRRKADAANAALVSGGVENGKRF
ESQCKSEKHPKRGVSPKMAODSDSGKRSKSGKMFEDLKKRGKQ"
complement(join(19162..19699,19767..19833,20044..20164,
20234..20378,20482..20623,20693..20738,20826..20877,
20965..21156))
/note="unknown protein: similar to ESNs gb|1741672.1,
gb|A19927710.1, and gb|AA586078.1"

CD5

join(13444..13814,13956..14310,14399..14662)
/note="similar to dihydrogenase protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.5"
/protein_id="AAAF71795.1"
/db_xref="GI:8052531"
/translation="MAVLSIKPVAIPKSGSLPIVIDMSDEPSKHALVAKADEGFEKVV
IHNGVSAELVSLHEHETVDFPSLSEKTYOYVAPFGNGVIGKRGNDVGWVEFLFN
ANDHQSQSPSLIKSPGTGRNALIEYTSVRKMTDGVLEKIRIDGIGLIPRNTLSL
VSDMDTSLIRLNHPPCPLSNKTKNGKAVNGICEHTDDPIISLVLSNNTSGQILN
NGSNIIVPBDITSEFFNVGDSLOYMTNGKRSVRRHRLAACKSRVSMYTFGSPSLT
QRIALPQLDINEDRLYEFTWSELYKSTYNSRLSDNRLOQFERKTIKLN"
complement(join(15755..16715,16810..16829))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F3f9.6"
/protein_id="AAAF71812.1"
/db_xref="GI:8052548"
/translation="MKPRGLIGRSELPOROSPRLTRSLSTSSDPHLISRPITDRSP
KGLDRSPRSRGPTDPLISOKIKLGRSGESLQIGAOEELIKTKOILAKAPAKR
AOEELHRRKSKSPNTPAPERDIPEDGHOETDVEVLDEKAKSEKTKNDLASKEDQ
INVLAKRLYDLERKSVLSSENETLKDQILKTDIVEMSCAKADELSKVSQIGEELE
ESNETTAKIKLESEVGEAKETLAEAKMKLVQVEGKRRKADAANAALVSGGVENGKRF
ESQCKSEKHPKRGVSPKMAODSDSGKRSKSGKMFEDLKKRGKQ"
complement(join(19162..19699,19767..19833,20044..20164,
20234..20378,20482..20623,20693..20738,20826..20877,
20965..21156))
/note="unknown protein: similar to ESNs gb|1741672.1,
gb|A19927710.1, and gb|AA586078.1"

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length
33987	AGAATCCAAATTTAAACCCCGGTGACAGACACCGCGCTCATCCCACTTAAACCGCGGATCTGCGG	34046	0	73.1%	245	95769
305	gtttcttcgtgtaagcagag---aagctaacccgctctctaaacggctatctctagt	361	0	77.1%	305	95769
362	ttcagatccgcggaatttgcctcggtggaatgaacacccgctcgggtatcttcgg	421	0	77.1%	362	95769
34107	TCCAAATAGCCGGGAACTTCTCCCGGTAACCGGAAACCTGTGTGATTAACCTTCCGG	34166	0	77.1%	34107	95769
422	tggtcgaaacttcgcgattccatcaaaagagtgatgtgctgcaacggaagctaacccac	481	0	77.1%	422	95769

Db	34167	TGACTGGAACAGTACCAAGATGCAATTCAGAGGATTTACTGCAAGAAACGGAGCAAAATCCAC	34226
Qy	482	ttcaacgaacccggtgacaaagttcaacacattcttcgaacgaacggtatggtttcaacccggtca	541
Db	34227	tttcacaaacacagctccggccacacatttttttcgacgcgaacacgatatgtaacacgcgtcc	34266
Qy	542	aattcgacaacggtttcaagcttaagcttaagcttcgcggttttactcaagctaacccggtttgtc	601
Db	34287	gggttcgataaaccgctcggtttagctacgcttgctggtttaccgaataaacccggttggttc	34346
Qy	602	aggaacgtaatltgggttcgaccggtttccccaaagccatcggttgagcttcaacggtccaca	661
Db	34347	aaagaaagtaaatgtgggtcgtccgggtttccccaaacaaatctggagcgtccacggacatt	34406
Qy	662	ccggtatgcccacacacagctatcttcaacgaagctcgaacggatataagttcaacccggt	721
Db	34407	tagaattgcccacaaagcttatgctcttcaataccocggcgctattttgggttatgctgaccca	34466
Qy	722	cacaacgaacccggttgaagctaaacccggttttgcttattcaatggccggttattggtcta	781
Db	34467	ccggaagacatcggtgctgcttaacgcgcggtttgattttcaatgatgctatcttttagcca	34522
Qy	782	ttctcgaaagatgatattactttacccaagttccgaatcccaatgtagatttaaaaacccg	841
Db	34527	tgctccaaacacgattttaccgtacacatgtaaaagttaactcaaacccgagattttagaacatt	34586
Qy	842	ttgctgctgctgaattttgtatgacaattagaatcccaatgatttgccaccggaaagctg	901
Db	34587	cgggttcggtacgatttttcgacgcgttaagtttaaatcaacaaatgatagccaccgaaatatg	34646
Qy	902	acccggaatccggttgaaacttcgcttctaagctaaagcttcaagacgttctaactaa	961
Db	34647	atcccgaaacccgagaaactcttcgttttaagctacgacgcgtttcaaaagccttactttaa	34706
Qy	962	aatacttccgattcttcaacggacggaactaaatacccggaagcttgaagattcaagcttgaac	1021
Db	34707	aaatcttccagatttcacatccgacggagagaaatcaccaacgcgcgaattccgcttgatgc	34766
Qy	1022	agcagaagctgatgcaacgatttcggaattcagaagaagacgttcgtcgtacacttgcaagc	1081
Db	34767	aaccacacgatgatccacgatttcgcgattactagaaattttcgtagtgatttccagaccac	34822
Qy	1082	aagctgttttcaagctcgcgcggaagatbaccgcggttggtctccggttttcaacaaaga	1141
Db	34827	aagtggtatttcgatttgccggagatataaagagtggttctccgggtgtttacgacgaca	34886
Qy	1142	acaaggtcgaagaatcgcggtatttagacaataacgcggaagattcatcgaaacattaagt	1201
Db	34887	acgaagaaatcaagatttcgggatttttaaaaaaacgcgaagatcgttcgctgattccaat	34948
Qy	1202	ggattgattcctcaagtttgcctgttcttcacatctcctggaagcgttgggaagacgaaga	1261
Db	34947	ggattccaaattaccagatttttttttttttttcattataggaacctttgggaagaaaccca	35006
Qy	1262	cagatbaagctgctcgatagaaggttcctgatatgactccacaacgaactaatttcaacga	1321
Db	35007	cagacgaggttgctcgatttgatctatgcatagcgcacctgatttaatttttcaacgaac	35066
Qy	1322	ctgacgaagaatctcaagatgttcctgtctgaaatccgcctgaaatcctcaaaacccggtga	1381
Db	35067	acgacgaataacatttcagagtggttttgcggagatagagcttaaacctttaaaccagcgga	35122
Qy	1382	caactgcgcgttcgcatcatcttccaaacgaagatccaagaatccaacccctggaacgagga	1441
Db	35127	caaacagtgtagacacgggttatcttc-----ggaaacatttttaatttcgaaagcgttatg	35177
Qy	1442	tcacagaacacatgctcgcgcgctgaacaaacgaattcgttacttggttttagccgaacg	1501
Db	35178	tcacacgggaattttattagttagaataaacccggatgtctttatcgtgcttttaacccaac	35237
Qy	1502	ggcttaaatcttcaagatttcgctaaagtgtgattctcaactcgaagaagttlaagaaacatc	1561

Db	35238	GGCCCTAAAGTGTCCCGGTTTCGCTTAAAGTGGAGCTTATTCGACCGGAGAAAGATTCGAAAGCTATA	35297
QY	1562	tttacgagcataacccgttacgagagagagagccctcgtttctccccggagagagagagag	1621
Db	35298	TTTTACGGAGAGGGGAAATATACGGAGGAGAGCCCTCTGTTCTACCTTCGCGTACGGAGAG	35357
QY	1622	aagacgaagagataactcctcgtttctccttaacagagaaacacatcggaatcgaggttac	1681
Db	35338	AAGACGAGAGGTTACTTAATGCTGTCTGTTACGACGAGAGAGAGAGTGAATGGAACTCC	35417
QY	1682	agatagttaacgcgttagcttagctagaaggttgaagcaacggttaacttcgctcaaggttc	1741
Db	35418	AACCTAATAACGCGCGTTATATAGAGCTTGAAGCTAACCGTAACTCCGTCGAGAGTGC	35477
QY	1742	cgtaaggaattacaggtatcatcatcagcagcagatgattggcgaagcagagtcgctgtg	1799
Db	35478	CTTACGGTTTCCACGGAACTTCATCAGTAAGGAAGATTTATCGAAGCAAGCTTTGTG	35535
RESULT	6		
LOCUS	AX148316	1818 bp	DNA linear PAT 08-JUN-2001
DEFINITION	Sequence 15 from Patent EP116794.		
ACCESSION	AX148316		
VERSION	AX148316.1	GI:14347203	
KEYWORDS			
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;		
	Lycopersicon.		
REFERENCE	1 (bases 1 to 1818)		
AUTHORS	Iuchi, S., Kobayashi, M. and Shinozaki, K.		
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene		
JOURNAL	patent: EP 116794-A 15 18-JUL-2001;		
FEATURES	Riken (JP)		
source	location/Qualifiers		
	1..1818		
	/organism="Lycopersicon esculentum"		
	/db_xref="taxon:4081"		
CDS	1..1818		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC41199.1"		
	/db_xref="GI:14347204"		
	/translation="MATTTSHATNTWIKTKLSPMSKDFEPAASISILKLNQHRKOSI		
	NINSLQAPLTHPKPQSNAYCPKRNITISHPKQENNNSSSSSKWLVYKAAAMAH		
	DAVSALTKHELEHPLPKTADPRVOISGNAPVPEINPCQSLPTGKTRPKVQCYVVA		
	NGANLPDEPTAGHFFEDBDGNVAHVORNGSASTCRFETETERLVOEKALDPVPFPA		
	IGELHGHGAIKMLFYARGLEGVADHSKGTGVANAGLVYNNRLAASEDDLPYHVK		
	VTLPGLDKTEGRFEDQLKSTMTALHPKLDVSGELFALSYDVLQPYLKYFRSKNGE		
	EKSNDVEIPVEDPTMMHDFATFENFVLPDQVVKMSKEMIRGSPVYDKNKYSRFG		
	ILDKYAKGSDLKKEVEDCFCHLMANMEAEDEIVISGKTPDSINEDDEGII		
	KSVSEIRKDKTKGSTRKSTIENDDEGVNLAEAMVNNKDKIGRTETAYALAIAPMP		
	VSGRAKVNLPTEVEKRTGKGNKKGGEPLTPPPNSKSEDDGIIIAVHDEKEMKS		
	LQIVNASKLEATVYKLPSPVYFGEHGFIVANDLANOA"		
BASE COUNT	569 a 383 c 369 g 497 t		
ORIGIN			
Query Match	45.8%	Score 824.6	DB 6; Length 1818;
Best Local Similarity	69.6%	Pred. No. 2.5e-234;	
Matches 1134; Conservative	0;	Mismatches 489;	Indels 6; Gaps 1;
QY	170	cctcagctcttattcctccttaagcaatcatcaactctccgccattgtgtttaaagcca	229
Db	185	CTTCAATATATCAACCAAGAAATATACATTTCACACCCCAAAACAGAAACAAACA	244
QY	230	aagccaagaatccacaacataaacagatgaattgttccagagagcgcgcgacgcgt	289
Db	245	ACTCCTCTTCTTCTTCACACTTCCAGTGTGAATTTAGTGCAGAAACACGACGCAATGCGCTT	304

QY	230	tggacgcgcgagaggttccctcgtcaagcagagaagctaacccgcttctctaaagc	345
Db	305	TAGATGCTGTGTAAGAAAGGCTTTAACTAAATGTAACATTGTAACACCCCTTGGCGAAACAG	364
QY	350	ctgatctcagtagtccaatctcgccggaattttgctcgg tgaatgaacgcgcgtccgc	409
Db	365	CCGACCCACAGTCCACATTTCTGGGAAATTTGCTCGGTAACGGAAATCCAGCTGTC	424
QY	410	gtaacttcgcgtgtgtcggaaaactcccgatctcaatcaagaagtgatgtgcgaag	465
Db	425	AATCTCTTCGGGTCCACCGGAAAAATACCMAATGTTGTAAGCGCTTTACGTTTCGAACGC	484
QY	470	gagctaccccacttcacagacgcggtgacagtgcaacactcttcagcagagaagctatg	529
Db	485	GAGCTAACCCCTTTTGTGAACCAACCGCCGACACCATTTTTCACGCGCACGSTATAG	544
QY	530	ttcacgcgcctcaaatctgaaacaggttcagctagctacgctctgcgcgtttactcagcta	589
Db	545	TTTACGCGCTCAATTTCAAAAATGGGCGGCTGTATGACTTGGCTTTCACTGAACAG	604
QY	590	accggttgttcaggaacgtccaattggtgtcgaacggttttcccgaagcaatcggtgagc	649
Db	605	AGAGCTGTGTCAAGAAAAGCTTTTGGGTGCCCTTTTTCCTTAAGCATTTGGTGAT	664
QY	650	ttcacgcgcacacccggtatgtcccgacatagtctatcttcacgcagagcgtgacgcgta	709
Db	665	TACATGGTCACTGTGGAATTTGCAAGGCTTATGCTGTTTACGCTGTGGGCTCTTGAGAC	724
QY	710	tagtcgaccgcggaacacgcgaacccggtgtagctaaegcgcggttgtctatctaatgyc	769
Db	725	TTGTGTGATCACAGTAAAGGAAGCTGTGTGCACCAACCCCGTTTAACTATTTCAATAAAC	784
QY	770	ggtatgtgcctatgtcggagagatgatatcttaacctaacagttcaagatactcccgaatgag	829
Db	785	GATTACTGTCATGTCCTGAAGATGATTTTGCTTACCATGTAAAGGTAAACCCACCGGCG	844
QY	830	atttaaaaaaccggtgtgcctcgtatcttgatltgagacattgaaatcaatcaatgattgcc	889
Db	845	ATCTTAAACAGAGGCTCGATTCGATTTTCACGCGCACGTAAATTCACCATGATACCTC	904
QY	890	accggaagtcgaccgcggaatccggtgaactcttcgctcttaagctcagagcgttcgttcaa	949
Db	905	ACCCAAACCTCGACCCAGTTTCCGGTGAAGTATTTCTCTTAGCTACGATGGATTTACA	964
QY	950	agccttaacctaaatlaacttcogaatttcacacgggaacgaataacacccgcgcgcga	1009
Db	965	AGCCATTACTCAACTACTTCAGATTTTCAAAAAATGGGAAAATCAATGATGTTGAAA	1024
QY	1010	ttcacgcttgatccagccaaagatgatgcaggaatttcgcgatcacagaagaacttcgtccg	1069
Db	1025	TTCCAGTTGGAAGACCCAAACATATGATGATGATTTCCCAATTTACTAGAACTTTCGTCGA	1084
QY	1070	taactgaccagaacgtcgttttcaaacgtgcggaagtgatccgcggtgtgctccggtg	1129
Db	1085	TTTCTGATCAACACAGTGTGTTTCAAGATGTCATAATGATCCGTGGAGGTTTACCGGTG	1144
QY	1130	tttacgacaagaacaaggttcgcgaagatccggatctttagacaataacgcgcgaatctac	1189
Db	1145	TTTTCGACAAACAAAGATTTCGCCGATTTGCTATTCGATTAAGATACGGAAAGATGGGT	1204
QY	1190	cgaacatlaagtgatgtatgtatgcacgaattgtcttcgctcatcatctcggaaacggttgg	1249
Db	1205	CTGATTTGAAATGGGTTGAAGTACCTGATTTGTTCTGTTTCCACCTCTGGAATGCTTTGG	1264
QY	1250	aagagccgaacacgatgaaagtcgttcgtgtataggttcctcgtlatgactccacaagactaa	1309
Db	1265	AAGAAGCAGAAACAGATGGAATGTTGTAATGTTGCATGTATGACACACACAGCTCCA	1324
QY	1310	tttcaacagagtcgtcgcgagaatctcaagaagtctcctgtctgtaaatccgcctgtaattca	1365
Db	1325	TTTTTCAAGATGTGATGAAGGCGCTAAAGAGTGTTTTATTCGAAATCCCTCTCAATTTGA	1384
QY	1370	aaacccggtgaatacactccgcctctccatatactccaaacaaatcaacaaatccaactcc	1429

Db	1385	AAACAGGGAATCAACAAGAAATTCATTAATGAAACCCCGATGACAAAGATTTAG	1444
Qy	1430	aaggaagatgttcaacaagaacatgtctgcgcgttaaaccaatctgtaactgtt	1489
Db	1445	AAGCGGAATGTAACCGAACAACCTCGAAGGAACAAGATATGCTTATTTGGCTA	1504
Qy	1490	tacgcgaacccgttgccctaaatctcaagattcgtcaaaagttcatctactactgagaag	1549
Db	1505	TCGCTGAACCATGGCCAAATTTCTGCTTTTGCAAAAGTAACCTGTTACACGGTGAAG	1564
Qy	1550	ttaaagaacatcttcaagcgataaccgttcaagagagacgtctgttctcccgag	1609
Db	1565	TTGAGAATTCATTATGTTGTAACAATATGTTGGGGAACCTCTTTTATACCAAGAG	1624
Qy	1610	-----aagggagagagaaagcgaagatcatctctctgttcttgcgttcaagcagaaga	1663
Db	1625	ACCCCAACAGAGGAACAAACGATGTTATATTTAGCTTTCGTACGATGAGAAAG	1684
Qy	1664	catggaatcgaagttacagatagttacgcgcgttagcttagaggttgaagcaagctta	1723
Db	1685	AATGGAATTCGAACGTGAATTTGTTAACGCAATGAGTTTGAAGTTGAGGCACTGTGA	1744
Qy	1724	aacttcgcgtcaagggttcgcgaatcttcacggtatctcatcgtgacgcgagattcgtg	1783
Db	1745	AGCTTCATCAAGAGTCTCTATGATTTCAATGMAACATTCATAACCCCAATGATTTGG	1804
Qy	1784	cgaagcag	1792
Db	1805	CAAAATCAGG	1813

RESULT	7
LENEOXANT	
LOCUS	2171 bp mRNA linear PLN 22-MAY-1998
DEFINITION	Lycopersicon esculentum mRNA for nine-cis-epoxycarotenoid dioxygenase.
ACCESSION	297215.1 GI:2769641
VERSION	9
KEYWORDS	nine-cis-epoxycarotenoid dioxygenase.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE	1 (bases 1 to 2171)
AUTHORS	Burridge,A., Griewe,T.M., Jackson,A., Thompson,A. and Taylor,I.B.
TITLE	Structure and expression of a cDNA encoding a putative neoxanthin cleavage enzyme (NCE) isolated from a wilt-related tomato (Lycopersicon esculentum Mill.) library
JOURNAL	J. Exp. Bot. 47, 2111-2112 (1997)
REFERENCE	2 (bases 1 to 2171)
AUTHORS	Burridge,A.
TITLE	Direct Submission
JOURNAL	Submitted (01-JUL-1997) Burridge A., The University of Nottingham, Physiology and Environmental Science, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK
REMARK	Revised by [3]
REFERENCE	3 (bases 1 to 2171)
AUTHORS	Burridge,A.
TITLE	Direct Submission
JOURNAL	Submitted (09-JAN-1998) Burridge A., The University of Nottingham, Physiology and Environmental Science, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, UK
COMMENT	On Jan 13, 1998 this sequence version replaced gi:2243153.
FEATURES	Location/Qualifiers
source	1..2171
organism	"Lycopersicon esculentum"
db_xref	"taxon:4081"
CDs	79..1896
product	"nine-cis-epoxycarotenoid dioxygenase"

Db	1385	AAACAGGGAATCAACAAGAAATTCATTAATGAAACCCCGATGACAAAGATTTAG	1444
Qy	1430	aaggaagatgttcaacaagaacatgtctgcgcgttaaaccaatctgtaactgtt	1489
Db	1445	AAGCGGAATGTAACCGAACAACCTCGAAGGAACAAGATATGCTTATTTGGCTA	1504
Qy	1490	tacgcgaacccgttgccctaaatctcaagattcgtcaaaagttcatctactactgagaag	1549
Db	1505	TCGCTGAACCATGGCCAAATTTCTGCTTTTGCAAAAGTAACCTGTTACACGGTGAAG	1564
Qy	1550	ttaaagaacatcttcaagcgataaccgttcaagagagacgtctgttctcccgag	1609
Db	1565	TTGAGAATTCATTATGTTGTAACAATATGTTGGGGAACCTCTTTTATACCAAGAG	1624
Qy	1610	-----aagggagagagaaagcgaagatcatctctctgttcttgcgttcaagcagaaga	1663
Db	1625	ACCCCAACAGAGGAACAAACGATGTTATATTTAGCTTTCGTACGATGAGAAAG	1684
Qy	1664	catggaatcgaagttacagatagttacgcgcgttagcttagaggttgaagcaagctta	1723
Db	1685	AATGGAATTCGAACGTGAATTTGTTAACGCAATGAGTTTGAAGTTGAGGCACTGTGA	1744
Qy	1724	aacttcgcgtcaagggttcgcgaatcttcacggtatctcatcgtgacgcgagattcgtg	1783
Db	1745	AGCTTCATCAAGAGTCTCTATGATTTCAATGMAACATTCATAACCCCAATGATTTGG	1804
Qy	1784	cgaagcag	1792
Db	1805	CAAAATCAGG	1813

Db	1385	AAACAGGGAATCAACAAGAAATTCATTAATGAAACCCCGATGACAAAGATTTAG	1444
Qy	1430	aaggaagatgttcaacaagaacatgtctgcgcgttaaaccaatctgtaactgtt	1489
Db	1445	AAGCGGAATGTAACCGAACAACCTCGAAGGAACAAGATATGCTTATTTGGCTA	1504
Qy	1490	tacgcgaacccgttgccctaaatctcaagattcgtcaaaagttcatctactactgagaag	1549
Db	1505	TCGCTGAACCATGGCCAAATTTCTGCTTTTGCAAAAGTAACCTGTTACACGGTGAAG	1564
Qy	1550	ttaaagaacatcttcaagcgataaccgttcaagagagacgtctgttctcccgag	1609
Db	1565	TTGAGAATTCATTATGTTGTAACAATATGTTGGGGAACCTCTTTTATACCAAGAG	1624
Qy	1610	-----aagggagagagaaagcgaagatcatctctctgttcttgcgttcaagcagaaga	1663
Db	1625	ACCCCAACAGAGGAACAAACGATGTTATATTTAGCTTTCGTACGATGAGAAAG	1684
Qy	1664	catggaatcgaagttacagatagttacgcgcgttagcttagaggttgaagcaagctta	1723
Db	1685	AATGGAATTCGAACGTGAATTTGTTAACGCAATGAGTTTGAAGTTGAGGCACTGTGA	1744
Qy	1724	aacttcgcgtcaagggttcgcgaatcttcacggtatctcatcgtgacgcgagattcgtg	1783
Db	1745	AGCTTCATCAAGAGTCTCTATGATTTCAATGMAACATTCATAACCCCAATGATTTGG	1804
Qy	1784	cgaagcag	1792
Db	1805	CAAAATCAGG	1813

BASE COUNT	688 a	435 c	431 g	617 t
ORIGIN				
Query Match	45.8%	Score 824.6;	DB 8;	Length 2171;
Best Local Similarity	69.6%	Pred. No. 2.6e-234;		
Matches 1134;	Conservative	0;	Mismatches 489;	Indels 6; Gaps 1;
Db	170	ctcagctcttcaatctccctaaagcaatcaacatctccgcacatgtgttgaagcca	229	/protein_id="CAB10168.1"
Qy	170	ctcagctcttcaatctccctaaagcaatcaacatctccgcacatgtgttgaagcca	229	/db_xref="GI:2769642"
Db	263	CTTCAATTTATCAACACACCAAGATTAATACAACTTCAACCCCAACAAAGAAACA	322	/translation="MATTTGATATTTATKTKLSMPSKSEFGPAASNLILKLNHROSL
Qy	220	aagccaagaatcccaacctaacaagatgaattgttccagagagcgcgcgcgcgcgcgc	289	NINSSIDAPPLIHPKROSSNTOTPKNNITISHPKRENNSSSSSTSKNNLYOKAAMAL
Db	323	ACTCTCTCTTCTTCAACTTCCAAAGTGAATTAATGATGTCGAGAAACACGAAATGCTT	382	DAVESALTKEHELEHLPKRTADPVOISGNATPEVENVCSLPVTKGIPVCQGVYR
Qy	290	tgaagc	349	NGANLPEPTGHHFFPDGDGVAVHVOVSKNSGASACRFEETETERLVOEKALGRPEYFA
Db	383	TAGATGCTGTGAAGAAGTGTGTTAACTAAACATGACACTTGAACACCCCTTCCGGAACAG	442	IGELHSGIARLMLFYARGLFGVDHSGKGTGVANACLVEFNNTLLAMEEDLPEYHVK
Qy	330	ctgactcctagtggttcaagatccgcggaatcttgcgcgcgcgcgcgcgcgcgcgcgc	409	VTPTGDKTREGRPDPGDLKSTMTAHKRLIDVSGELPALSDYLOKPYLKYFRSKNG
Db	443	CCGACCCACGAGTCCAGATTTCTGGGAATTTGCTCCGGTACCGGAAATACCAAGTCTGTC	502	EKSNDVETPVEDPTMMDFAATEHFPVYIPDOGVFKSEMTRGSSPYVYXNKYSREG
Qy	410	gtaactcctcgc	469	ILDKYANDGSDLKWVEVDCFCFLHMANWEAEATDEIVTGSCTPPDSITNEDEBL
Db	503	AATCTCTTCCGGTCAACCGGAAATTAACCAATGCTTCAAGGCGCTTTCGTAACACG	562	KSVLSEIRLMLKTKSSTKRSKSLIENPDQVNEAEAMVNRNKLGRRTXYALAIAPWK
Qy	470	gagctaacccaatccagc	529	VSGPAKYNLFTGEVEKEFYIGDNKGGEPLFPRPNKSEEDDGEILAFVHDEKEMSE
Db	563	GAGCTAACCCCTCTTTTGAACCAACCGCGGACACCATTTCTTCGACGGCGACGATAGG	622	LQIYNASLTKLEATVTKLPSRVYGFHFIFINANDLANOA"
Qy	530	ttcagc	589	
Db	623	TTACATGCTGCTGATTAATTAATAATGAGTGGCTAGTTAGCTTCCGCTTCACTGAACAG	682	
Qy	590	accggttgttcaagaaatgaatgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	649	
Db	683	AGAGGCTTGTTCAGCAAAAAGCTTTGGTGGCCGCTGTTTCCCTAAGCCATGTGTAAT	742	
Qy	650	ttcagc	709	
Db	743	TACATGCTGCTGCTGATTAATTAATAATGAGTGGCTAGTTAGCTTCCGCTTCACTGAACAG	802	
Qy	710	tagtcgaacccgc	769	
Db	803	TTGTTGATTCACAGTAAGAGACCTGCTGTGGCAAAAGCCGCTTACCTAATTTCAATTAAC	862	
Qy	770	gttatttgctatctcgaagagatgaattacacttaccaggttcaagtcactcccaatgag	829	
Db	863	GATTACTGCTGATGCTGAGATGATTTGCCCTTACCAATGTAAGGTAAACCCACCGGCG	922	
Qy	830	atlttaaacccgttgctcgc	889	
Db	923	ATCTTAATAAACAAGAGGTGATTCGATTTTCGAGGCGACAGTAAATTCACCAATGATAGCTC	982	
Qy	890	accggaagctcgaacccgc	949	
Db	983	ACCCAAAGCTCGACCCAGTTTCCGGTGAAGCTATTGCTCTTATGTCAGATGTGATTCAGA	1042	

Db	758	GGCTATGCTGTTTACGCTGTGGCTCTTCGACCTATTGATCAAGTAGAGAACTG	817
Qy	734	gtgtagctaaagccggttggctatctcaatgcccgttattgctctgcccgaagatg	793
Db	818	GAGTTGGCAAGCGCGTTTGTAGTCAATTAACCGGATCTTCTGTCGTAAGATG	877
Qy	794	attaccctaccagttacagatcccaatggagatttaaaacccgttgctgctcg	853
Db	878	ATTTGCTTACCAATGTAAGTAACACCCACCGGATCTTAAACAGAGGCTCATTCG	937
Qy	854	atttgaatggaacaattagaaatccacaatgatgtccaccaccgaaagtcgaccgaatcgcg	913
Db	938	ATTTCGAGCGCAGCTAAATCCACCATGATAGCTCACCCAAAGCTCGACCCAGTTCCG	997
Qy	914	gtgaactctcgctttaaagctgaagcgtcttcaagaccttaactaaataactcgat	973
Db	998	GTGACCTATTGCTCTTGTGCTACGATGTGATTCAGAACCCATACCTCAAGTATTTCAGAT	1057
Qy	974	tctaccggaagcaataatcaccggaagctcgagatcagcttgatcagaaccaagatga	1033
Db	1058	TTTCAAAAATGGGAAAAATCAATGATTTGAAATCCCACTTGAAGACCAACATGA	1117
Qy	1034	tgacagatctgcgagttacagagaactcgtctgctacgtacccaagcaagctgttca	1093
Db	1118	TGCATGATTTGCAATTAAGTACGGAATTCGTCATCTCATTCGATCAACAGCGTTTCA	1177
Qy	1094	agctgcggaatgatacccggttggtctccggtggtttaaagaagaagaagtcgcaa	1153
Db	1178	AGATGCTGAAATGATTCGCGTGTGACCGGTGTTTACACAAAGAAAGTGTCC	1237
Qy	1154	gattcggatctttaaacaataacgcggaagattcagaacataagtcgattgctc	1213
Db	1238	GATTGGGATTTCTGATTAAGTACGCGAAAGATGGCTGATTTGAAATGGTTGAATTC	1297
Qy	1214	cagatgtctcgtctcattcattctcgtgaagcgttggaaagaccagaacagatgaagtc	1273
Db	1298	CTGATTTGTTCTGTTCCATCTCTGGAATGCTTGGGAAGAACACAGAACATGAAATCG	1357
Qy	1274	tcgtgatagggctcgtctgctgactccacagactcaatttcaagagctcgcgaagatc	1333
Db	1358	TTGTAAATGTTTCATGATGATGACACACAGACTCCATTTTCAATTAATGTGATGAGGAC	1417
Qy	1334	tcaagagctcgtcgtcctaaatccgcctgaatctcaaaacccgtgaatcactgcgctc	1393
Db	1418	TAAAAAGCGTTTATCCGAATTCGTCATTTAAAAACAGGGAATTCACAAAGAAAAG	1477
Qy	1394	cgatcattccacaagaagatcaacaacactcgaagcgaagatggttcaacagaaca	1453
Db	1478	CCATTAATGAAACCCGGATGAAACAAATTTAGAGGCTGGAATGGTGAACAGAAACA	1537
Qy	1454	tgctgcgcgttaaaacccaatcgttacttggtctttagccgagccggtgcttaagct	1513
Db	1538	AACCTTGAAGGAAACACAGTATGCTTATTTGGCTATTGCTGAACCATGCGCAAAAGTTT	1597
Qy	1514	cagagatcgtcaaaagtatcactactgagaagttlaagaacatctttagccgata	1573
Db	1598	CTGGTTTGGCAAAAGTAAACCTGTTTACAGGTGAAGATTGAGAAATTCATTTTGGTGACA	1657
Qy	1574	accgltacggagagagcctctgttctcccgagag-----aaggaagaggaagagag	1627
Db	1658	ACAAATATGAGGAGGAGCGCTCTTTTAAACCAAGAGACCAACAAAGGAGGAAGCG	1717
Qy	1638	aaggtataactcctcgttctgctcagcagagaagaacatgtaaatcgcgaattacagatg	1687
Db	1718	ATGGTTATATTTTACCTTCTGTTGATGATGAGAAAGAAATGACATCAGACTTCAAAATG	1777
Qy	1688	ttaagccgttaagcttaagagttgaagcaaggttlaaacttcgccaaggttccgttacg	1747
Db	1778	TTAAATGCAATGACTTTAAAGTTAGAGGCAACAGTGAACCTTCATCAGAGACTTCTTATG	1837
Qy	1748	gatttcaagatcattcattcggagccgaatgatttggcgaaagcag 1792	

Db	1838	GTTTTCATGGAACCTTCAATTAATGCAATGATTTGGCAATCAGG	1882
RESULT	9		
AXI48312			
LOCUS	AXI48312	1839 bp	DNA
DEFINITION	Sequence 11 from Patent EP116794.		linear
ACCESSION	AXI48312		
VERSION	AXI48312.1	GI:14347199	
KEYWORDS			
SOURCE			
ORGANISM			
	Vigna unguiculata		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;		
	Vigna.		
REFERENCE	1 (bases 1 to 1839)		
AUTHORS	Iuchi, S., Kobayashi, M. and Shinozaki, K.		
TITLE	Transgenic plants carrying neoxanthin cleavage enzyme gene		
JOURNAL	Patent: EP 116794-A 11 18-JUL-2001;		
FEATURES			
source	Location/Qualifiers		
	1..1839		
	/organism="Vigna unguiculata"		
	/db_xref="taxon:3917"		
	1..1839		
	/note="unnamed protein product"		
	/codon_start=1		
	/protein_id="CAC41197.1"		
	/db_xref="gi:14347200"		
	/translation="MPSASNTWPNATLPSPPFDLPSTSSPTNLDLRKSSNTIT		
	CSLOLHPKROYPTSTSTPATTTTPPIKTTITTTTTPRETNPLSDVQIPQK		
	NFLQKATADIVETALVSHERKRLPDTDPVQALAGNAPVPEHAADQGLVYVK		
	IPKCTIDVYVNGANPLPEPVAGHHFPEDGDMVNAKFTNCAASYACRFTTQLSOE		
	KSLGRVPFPAIGELHGHSGIARLLFTARLFLVLDVDSQGMVANGLVFNHLLA		
	MSEDDLPYHVRITPNQDITVGRVDFNQLSTMIAPKLDPVGDJHALSYDIQK		
	YLKFRSPDVKSPDVEIPLKEPTMHMDFATFNFVVPDQVYFRLKTEMITGSPV		
	YVDKNTSRFGILKHNKADANAMWIDAPDCEFLMVAAMEPEIEVYVIGSCMTPA		
	DSIFNECESIKSVLSERLNLPTGKSTRRIISDAQVNLAEAGVNRNKGRTQRA		
	YLALAEPRKVSGRFAKVDLSGEVKKHYGEEKRGGERLPLPNQKEDDGIILAFVD		
	EKEWSEIQIYNAONLKLKLEASIKLPSRPVIFGHGTFFHSKDLRQA"		
BASE COUNT	447 a 595 c 442 g 355 t		
ORIGIN			
Query Match	44.4%; Score 798.4; DB 6; Length 1839;		
Best Local Similarity	70.4%; Pred. No. 1.6e-226;		
Matches 1098; Conservative	0; Mismatches 456; Indels 6; Gaps 2;		
Qy	233	ccaagaatccacaacatacagatgaatttgcacagagacgagccgagcagcttgg	292
Db	281	CCAACCAACCAATACCTTCAAAAATGGAACCTTCTCAGAAACCCCTGCCATGGCTTGG	340
Qy	293	acgcgcgagaggggttctctgtgcacgcagagaagctacaccgcttctcaaacgctg	352
Db	341	ACCTGCGTCAAGCGCGCTCGTTCGACGACGCAACACCCGCTCCCAAAACGCGCG	400
Qy	333	atccatagttcagatgcgcggaatttgcctcggttgaatgaagaagccgctcggcgta	412
Db	401	ACCCAGAGGCTCCMAATCCGCGGAACTTCGCGCGCTGCCGAGCATGCGCGGATCAG	460
Qy	413	atcttcgagtgtaggaagaactccgatctcaataaagagatgfatatgagcaacgag	472
Db	461	GACCTCCCGGTGTCGGAATAATCCCAATGATGATGACGCGCTGTACGTGCCCAACGGTG	520
Qy	473	ctaaccacttcaagcgcggtgacaggttcaaccacttcttcgacggaagagttatgcttc	532
Db	521	CCAATCCGCTTACAGAGCCTGTGGCCGACCACTTCTTTCACAGCGGACGCAATGGTCC	580
Qy	533	acgcgctcaaatcgaacacggtttagctagctacgcttgcggtttactacgaactaac	592
Db	581	ACGCGGTGAAGTTCAAGAACGCGCGCGCCAGCTACGCTCGCGCTTCAACCGAGACGACG	640

QY	593	ggtctgtctcgaagacgcgaatcttggttcgacgggtttttttccaaagacgaatcggltgacgtc	652
Db	641	gctcttcgcgaagaaatctcttaagcccccgggtgtttccgaaaggccatcgggagacgtcc	700
QY	653	acggccaacacgggcatctgcccagactcatgtatcttaacgccaagatcgtacgcggtatag	712
Db	701	acggccatctccggcatcgccggcgctcctctcttcttaccggcgcggtctcttcggggctcg	760
QY	713	tcgaccgcggcaacgcggaacccgltgtagctcaacgcgcggttttgltctatttcaatgycggt	772
Db	761	ttgatgggtccccaaggcatctggcgctggcgcaaacccggtctcgtcttcaactcaacacccac	820
QY	773	tattggtcatctgtaggaatgatatttcaactccaagtccaagtccaactccaatggaagt	832
Db	821	ttctggcccatgtctcgaagacgattttacccttaccacgtgaattacaccccttaacggcgact	880
QY	833	taaaacacggtctggtcgggtctgattttgatgtgaacattgaatccaaatgattgtccac	892
Db	881	ttaccaccggttgccgcttttagactttcaacgggagctcaactcaacatgatgtccacc	940
QY	893	cgaagctgcacccggaaatccggtgaactctctgcctttaagctcaagcgtctgtttcaaaag	952
Db	941	cgaaatctggaccctcgtagagggagacctccagcgcgtcagctacagacgtcattcgaagac	1000
QY	953	cttactcaaatctcccgattctcaacgcggaaggaactaaatcaacccgacgtcgaattc	1012
Db	1001	ctttactctcaacttacttccgttt	1060
QY	1013	agctgtacagccaagatgatgacgattctgcgttcaagagaactctgcgtcgtac	1072
Db	1061	ccctgaagaccccaacatgatgatgacatttgcctaaacgagaaatttctgctgtctcc	1120
QY	1073	ctgacagcaagctgctttcaagctgctgcggagatgaatccggtgtgctccggtgtt	1132
Db	1121	ccgaccacacagctggtcttcaactaaacgagatgatcaacccggcggtctcccggtgtct	1180
QY	1133	acgacaagaacaaggtctgcgaagttcgggattttagacaataacgcgcgaagatctatcga	1192
Db	1181	acgacaaacaaacaaacccctcacgggttgggtttctgcacaaagatgggaagacgcgaatg	1240
QY	1193	acattaaatgtatgtatgctccagattgctctgctccatctcggaaacggtttggaag	1252
Db	1241	cgatccggtgatgtgacgcccggattgtttcttcttccactcttggaacggctggagcg	1300
QY	1253	agccgaacacagatgaagctcgtctgataggtgcctgtatgactccacccaagactaat	1312
Db	1301	agccgaacacagagagaggttggtgatgggtcgtcgaatgaacccctggcactccattt	1360
QY	1313	tcacagagctctgacggaatctcaagaagtctctgtctcgtcaaatccgctgaatctcaaaa	1372
Db	1361	ttcaacgaatttggaagaaagatttgaagcctcgtctcaagatgaagcgaacttgagaga	1420
QY	1373	ccggtgaatacaactcgcgcgtccgatactctccaaagagaatacaaaagtcacaactccgag	1432
Db	1421	ccggcaactccactcgcgcgcgccttatttctcgacccgcg -- aacaaaggaactcggtag	1477
QY	1433	caaggatgtgcaacaagaacaatctcgcgcgttaaaacaaactcgttacttgcgttttag	1492
Db	1478	ccggcatgtgtaacagaaacagactcggaaggaagaaacccagtttcgggtatctgcttgcg	1537
QY	1493	ccgagcgtgtgcctaaagtctcaagatctgactaagtgtatctcaactactctgagaagta	1552
Db	1538	cgagaccccttgcccaaaagcttccggcgttttcgaaatgtattgttcgagatgggaaagta	1597
QY	1553	agaacacttlaaggcgataacggttlaagggaagagactctgtttctcccggaagag	1612
Db	1598	agaaatcatattatggaagaaagaaagttcgggtggagacctgtgttcttctcc -- aacg	1654
QY	1613	gaaggaggaagaaagaaatacactctgttcttcaacagagaagaatgaaat	1672
Db	1655	gccaaaaaagaagacgatgggtatatttcttgscattctgtcgaacacgaaagaaatgaaat	1714
QY	1673	cgaggtacagatagttaacgcgcttagcttagagltgaagcaacggttlaacttcgt	1732

[illegible]

QY	353	atccctagctgtcsgatcgcgcgaatcttctgcggtgaaatgaacagccgcgtccgcyta	412
Db	526	ACCCGAAGGGTCCAAATATCGCCGGAACTTCGCGCGGATGAGCATGCCGGATCAAG	585
QY	413	atcttcgcgtgtgtcggaaaaactcccgatctcatcaagaagatgtatgtgtgcgaacgag	472
Db	586	GACATCCGGGTGTGCGAAAAATCCCAATATGATTAACGCGCGTGTACGCGCAACGGGTG	645
QY	473	ctaaccaactctaacgagccggttgaaaggtgcacaactctcttgacgagagatgtgtc	532
Db	646	CCAATCCGGCTCTTACGACCCCTGTGGCCGGGACACACTCTTCTTCAACGGCCAGCGCATGTCTC	705
QY	533	acgcgcgtcaaatctcgaaacacggttcagctagtacgctgtccggttactctcgaactaac	592
Db	706	ACCGCGTGAAGTTTACGAAAGGCGCGCCGACAGCTACGCTTCGCGCTTCAACCGAGACGAC	765
QY	593	gatttgttcaaggaaactcaatttgggttcgacccggtttccccaagccatcgtgtgcttc	652
Db	766	GTCCTCTCGAGGAAATCTCTAGGCGCGCGGTGTTCCGGAAGGCATCGGGGAGCTCC	825
QY	653	acgcgcacacccggttatgtcccgcatactatcttaacgcagagctgaacggttatag	712
Db	826	ACGGGCACCTCCGGCATCGCGCGGCTCTCTCTTACCGCGCGGTCTCTTCGGGGCTCG	885
QY	713	tcgaaccgcgcacacggaaccggtgtgtaagctaacgcggtttgttctatttcaatggccggt	772
Db	886	TTGATGGGTCTCCAGAGGCGATGGCGGTGGCGACGCGCGTCTGTACTTCAACAMCACC	945
QY	773	tattgtctatgtcggagagatgtattaccttaccaaagttcaagatctcccaatgagat	832
Db	946	TCCTGGGCATGTCCGAAGAGAGATTACCTTACACGTGAATTCACCCCTAACGGCGACT	1005
QY	833	taaaacacgttgtctgtctgcgatatttgatggaaatattgaatccaaatgattgtccacc	892
Db	1006	TAAACACCGCTTGGGCCCTTACGACTTTAACGGGAGCTTCAACTCAACAAATGATCGCCACC	1065
QY	893	cgaagctcgaccgcggaatccggtgaactcttcgctttaagctacagcgttcgttccaagc	952
Db	1066	CGAATCTGGACCCCGGTGACGGGAGCTCCAGCGCGTCACTACGACGTCAATTCAGAAAGC	1125
QY	953	cttaccctaaatattcttcgatcttccacgcggaacctaatacccggaacgtccgagatc	1012
Db	1126	CTTTCCTCAAGTACTCTCCGTTTCTCCCGGACGCGCTCAAGTCCCGCGAGTGAATTC	1185
QY	1013	agcttgtacgaacaaagatgatacgaattttcgagatcagaagaactctcgtccgtctac	1072
Db	1186	CCCTGAAGGAGCCCACTATATACAGATTTCGCCCTAAACGAGAAATTTCTGTCTGCTCC	1245
QY	1073	ctgaaccagcagtcgtcttccaagctctccgagagatgataccgcggtgtgtctccggtgtt	1132
Db	1246	CCGACCGACGAGGTGTCTTCAAACTAACGGAATGATGATCACCGCGGGGTCCCGCGTGTCT	1305
QY	1133	acgaacaagaacaaggttcgcaagatccgggaattttagaacaatacgcgcgaagaattcatga	1192
Db	1306	ACGCAAGAACAAAACTCTACCGGTTTGGGATCTTGACACAAGATCGAAGGACGCGAATG	1365
QY	1193	acattaaagtgtattgatgttcgcgcgaagattgtcttcgtctcatctatcttgaaagcgtttgag	1252
Db	1366	CGATGCGGTGATGACGCGCCGCGATTGTTCTGTCTCAACTCTTGGAACCGGTGGGAGG	1425
QY	1253	agccagaacaagaatgaagttcgtctgtatagaggttcctgtatagtactccacacgaactcaatt	1312
Db	1426	AGCCCGAAGACGAGAGGTGTGGTATTTGGGTCTCGATGATGACCCCTGGGACTCTCATTT	1485
QY	1313	tcgaagagcttcgacgagaatctcaagagtgtccgtctgtaaatccgcgcctgaaatctcaaa	1372
Db	1486	TCAACGATGCGAGGAGATGTTGAAGAAGCGTCTGTCAAGATTAAGGCTGAACCTTAGAGA	1545
QY	1373	ccggttgataactactcgcgcgtccgatactatctccaagaagatcaacaagtcaactccgag	1433
Db	1546	CCGGCAAGTCCACTCTCGCGCCCATTTATCTCCACGCCG---AACAAAGTAACGTGGAAG	1602

OY	1433	caggatggtcacaaggaacatgctcgccgcgttaaacccaattcgcttacttgcttag	1432
OY	1433		
Db	1603	CCGGCATGTGTAACGAACAAGCTCGGAAGAAGAACCCAGTTCGGATCTGGCTCTGG	1662
OY	1493	ccgaaccgttgacctaaagtctcaagatttcgctaagaattgatctactactggagaagta	1552
OY	1493		
Db	1663	CGGACCCTTGCCCCAAAGTCTCGGGCTTTGGCAAGATTGATTTGCTGAGTGGGGAAGTGA	1722
OY	1553	agaacacatttacgcgcatataccgtltiacggaagagaagccctcgtttlctcccgcggaagaa	1612
OY	1553		
Db	1723	AGAACTACATGATGATGAGAAGAAGATTCCGGGGGAGCCTCTGTTCTTCCC--AAGC	1779
OY	1613	gagggaggaagaaagcaagagatacatcctctcgttttcgttcaagagaagagaacatggaat	1672
OY	1613		
Db	1780	GCCAAAAGGAAGACATGGGATATATCTTTCGGCATTCGTGCACGACAGCAAGAAATAATGGAAT	1839
OY	1673	cggagttcacagatagttiaaccgccgttagcttagagagttgaagcaacaggttiaactccgt	1732
OY	1673		
Db	1840	CCGACTCTCAGATTGTGAATGCCCAAATTTTAAAGCTGGAAAGCTTCATCCAACCTCCCT	1899
OY	1733	caaaggttcgcgttacggaattcaacggttatcatcatcgaagccgatgattggcgaagcaag	1792
OY	1733		
Db	1900	CTCGGTTCCCTACAGGTTTTATGGAAGTTTCATTCATTCACAGGATTTGAGGAACAAG	1959
RESULT	11		
AF190462			
LOCUS	AF190462	2398 bp	mRNA linear PLN 19-JAN-2000
DEFINITION	Phaseolus vulgaris 9'-cis-epoxycarotenoid dioxygenase (NCED1) mRNA,		
ACCESSION	complete cds.		
VERSION	AF190462		
KEYWORDS	AF190462.1 GI:6715256		
SOURCE			
ORGANISM	Phaseolus vulgaris.		
	Phaseolus vulgaris.		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;		
	Phaseolus.		
REFERENCE	1 (bases 1 to 2398)		
AUTHORS	Qin,X. and zeevaart,J.A.		
TITLE	The 9'-cis-epoxycarotenoid cleavage reaction is the key regulatory		
JOURNAL	step of abscisic acid biosynthesis in water-stressed bean		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (26), 15354-15361 (1999)		
PUBMED	20079657		
REFERENCE	10611388		
AUTHORS	2 (bases 1 to 2398)		
TITLE	Qin,X. and Zeevaart,J.A.D.		
JOURNAL	Direct Submission		
STATE	Submitted (27-SEP-1999) MSU-DOE Plant Research Laboratory, Michigan		
UNIVERSITY	State University, Wilson Str., East Lansing, MI 48824, USA		
FEATURES			
Source	location/qualifiers		
	1..2398		
	/organism="Phaseolus vulgaris"		
	/cultivar="Top Crop"		
	/db_xref="taxon:3885"		
gene	1..2398		
	/gene="NCED1"		
	124..1971		
	/gene="NCED1"		
	/function="catalyzes the oxidative cleavage of the C-11,12		
	double bond of 9'-cis-epoxycarotenoids (c40) to yield		
	xanthoxin (C15) and C25-apo-aldehydes"		
	/codon_start=1		
	/product="9'-cis-epoxycarotenoid dioxygenase"		
	/protein_id="AAF26356.1"		
	/db_xref="GI:6715257"		
	/translation="MPSASNWTWINTLIPSSCSPPFKDLASTSSPTLLPFKRSS		
	NTNITSLTIQHYPKOYOPSTSTTPPIKRTTTTTTTTTPRKRLSDRKPP		
	OKNFLOKAATGTGDVETALVSHESHNP.LPKTDPRVOQLGNFAPYPERAADOLFV		
	VGR.LPKCIDGVYVRNGANPLYLEPAVGHHFEFGDGMVAIVFTNGSAVYACRFETOL		
	AOESKSPVEPKRAIGELHSGIARLLIFEARLFOLDVDSHGMSGVANGAEVFNNH		
	LIAASEDLPHRYITNSGDLTVRGDENGLNSTAIHPKILDPVNGDIHALSYDVV		
CDS			

OKPVIKIFRESAGDVCSPDVIDIPLKEPMMWMDFAITFNFYVVPDOOVYITLITINGG
SPVYDDKNKTSRBCGILDNKADNANMAMWIDAPBECFCCHLINAAWPEPTDEIVIGSGM
TPBDSINREDCDESLKSVLSEIRNLURKQSTRRIIISDAEYVNLAEKAMVNNKRIKGRIT
QFALALIAEPMPIKVSFGFAKVDIFSEGVYVYVYGEKRGDEPLFIPNDEEBRDGYILAF
VHDKRESELOIYNAONLKLEASIKLPSRVPYVGFHGTFSKDLRQKA"
627 a 687 c 516 g 568 t

[illegible]

QY	1127	tggttaagacacaagaacaaggtcgcgaagattcgggattttaagacaataatgcccggaatt	1186
Db	1307	TGGTTACGACACAGAAACAAACCTACGGTTTGGGATTTTGGATGAAGATGGGAAGGAGC	1366
QY	1187	catcgacaactaagtggatttga tgcctccagattgctctcgtctccatctctgaaagctt	1246
Db	1367	CGAATGGCATGCGGTGGATCGAAGCGGCGGAGATGTTTCTGCTTCCATCTCTGGAAAGCGT	1426
QY	1247	gggaagaccagaacaacagatgaagtcgctcgtgataaggttcctgtatagctccacagact	1306
Db	1427	GGGAGAGAGCTTAACACGACGAGATTGTGGTATTGGTGGGCTCCGATAGACCCCTGGGAGT	1486
QY	1307	caatttcaagagctcgtcaagaaatctcaagaagtgctctgctctgaatctcgctgaatc	1366
Db	1487	CCATTTTTCATATATGCGATGAGAGTTTGAAAGCGCTTGTGTCTGATGATAGGCTGAACT	1546
QY	1367	tcaaaacccggtgaatccaatcgcgtccgatacatctccaacgaagaatcaacaagtcaacc	1426
Db	1547	TGAGAGACGGGCAAGTCGACTCGCCGCCCATTTATCTCCGACGCCGC--AACAAGTGAAC	1603
QY	1427	tccaagaccggatggtccaacagaacaatgctgcgcgtcaaaaacaaattcgcttaactgg	1486
Db	1604	TTGAAGCCGGCATGTGTGAACACAAACAACTAGGAAGAAACTCATGTTGCTTATCTGG	1663
QY	1487	ctttagccgagccggtgcccctaaagctcaagatctcgctaaagtctgatactcaactagag	1546
Db	1664	CTGTGGCCGACGCCCTGGCCCAAAAGTCTGGGCTTTCCGAAAGTTGATTGTTCAGAGGGG	1723
QY	1547	aagctaaagaacaatcttctacgycgataacgcttlaagggagaaagcctcgtttctcccg	1606
Db	1724	AAGTCCAGATGATCATGTTGTGGAGAAAGAAAGTTGGTGGTGGGAGGCTCTGTTTCTTCCC-	1782
QY	1607	gagaagaggagaaagaaagacgaagatacatcctctcgttctcgtctcaagacgaagaacat	1666
Db	1783	--AACGGCGAAGAAAGAGGTGATGGGTATTTCTGGCATTCGTGCACGACGAGAAAGAT	1840
QY	1667	ggaatctggagttacaagataagttaaacgcgcttaagcttagaggttgaagcaacggttaaac	1726
Db	1841	GGAAATCCGAGCTGAGATTGTGAATGCCCAAAATTTAAAGCTCGAAGCTTCCATCAAC	1900
QY	1727	tcccgctcaaggtctcgttcaagattcaacgttacatctcaacggaacgtatctggcga	1786
Db	1901	TCCCTCTCGGCTTCTCTTATGTTTTCATAGGACACTTTTATTCATTCTAAGATTTTGAGGA	1960
QY	1787	agcaggt 1792	
Db	1961	AGCAAG 1966	

AC074176/c	AC074176	72058 bp	DNA	1 linear	PLN 19-JAN-2001
LOCUS	AC074176	72058 bp	DNA	1 linear	PLN 19-JAN-2001
DEFINITION	Arabidopsis thaliana chromosome 1 BAC T2H7 complete sequence.				
ACCESSION	AC074176				
VERSION	AC074176.5	GI:12321625			
KEYWORDS	HTG.				
SOURCE	thale cress.				
ORGANISM	Arabidopsis thaliana				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	1 (bases 1 to 72058)				
AUTHORS	Lin, X., Kaul, S., Town, C.D., Beato, M.-I., Creasy, T.H., Haas, B.J., Wu, D., Maitl, R., Romling, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Bartshead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.				
	Arabidopsis thaliana chromosome 1 BAC T2H7 genomic sequence unpublished				
TITLE	2 (bases 1 to 72058)				
JOURNAL	Town, C.D. and Kaul, S.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (15-JUL-2000)				
TITLE	The Institute for Genomic Research, 9712				
JOURNAL	Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org				

REFERENCE 3 (bases 1 to 72058)
AUTHORS Town,C.D. and Kaul,S.
TITLE Direct Submission
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
On Jan 19, 2001 this sequence version replaced gi:12280878.
Address all correspondence to:atetigr.org

COMMENT

BAC clone T2H7 is from Arabidopsis thaliana chromosome 1
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of several methods: Gene
prediction programs including Genscan+ (Chris Burge,
http://CCR-081.mt.edu/GENSCAN.html), GenemarkHMM (Mark Borodovsky,
http://genemark.biology.gatech.edu/genemark/), Glimmer4 (a variant
of Glimmer3, see Mihaela Pertea,
http://www.tigr.org/softlab/glimmer_hmm/glimmer.html, and
GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
mpertea@tigr.org), searches of the complete sequence against a
peptide database and the plant EST database at TIGR
(http://www.tigr.org/tdb/cgi.shtml). Annotated genes are named to
indicate the level of evidence for their annotation. Genes with
similarity to other proteins are named after the database hits.
Genes without significant peptide similarity but with EST
similarity are named as unknown proteins. Genes without protein
or EST similarity, that are predicted by more than two gene
prediction programs over most of their length are annotated as
hypothetical proteins. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
Simple repeats are identified by repeatmasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers

FEATURES

source

1..72058
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="I"
/clone="T2H7"

misc_feature

1
/note="28797 nt beyond this point were not included in the
submitted sequence due to an overlap with another BAC
(TIP2)"

mRNA

complement(<729..>1868)
/gene="T2H7.1"
complement(729..1868)

gene

/gene="T2H7.1"
/note="similar to hypothetical protein GI:2832643 from
(Arabidopsis thaliana)"
complement(729..1868)
/gene="T2H7.1"
/codon_start=1

CDS

/product="hypothetical protein"
/protein_id="AAG50853.1"
/db_xref="GI:12321627"

/translation="MSYLRSDDPVSRTHPEPOSTLSPDHFDLLPDSLLLLTFDVKVADY
KDLGRCCIVSRFHSVLVFNVCVISDDSSSDENHRELSLTATISDAGG
GGSFALPLVAFAPLFIKPFOMIGOLIGKRSSSDASFSAINDEIGVTHSPVOVAK
MGGRIRIKLEIPRGLGIEDGILLKRRADGSLDNCMITGASVIOSSNVKHENS
VDEDNGINPESFTNGGKLRVNTISLNASRHYLQPTINEHKSLDRLVSDAD
GGVGLCKNRQLEELRVTPLSASSASAKRTLVAPALNMRMLWVPELDLPDGTVLKATTV
AIRPSESKKEVCDAWSLDAEPEEPFGVAKMLIRRYCLEMSNF"
complement(12328..2392)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

mRNA

complement(<7213..>8049)
/gene="T2H7.2"

gene

complement(7213..8049)
/gene="T2H7.2"
/note="predicted by genemark hmm"
complement(7213..8049)
/gene="T2H7.2"
/codon_start=1

CDS

/product="hypothetical protein"
/protein_id="AAG50854.1"
/db_xref="GI:12321628"
/translation="MTEPMSYMIENPKFEPRKRRYSSMLTIFLPTITMIFHVE
VLSVSEFDKVLRFISNTLLILIAADGSFSDKESQDFIGETVVAATMRNDNS
PIPVLYRENTKDEIKNPKDVERNEDEDEPVKDIQVSPPEKIVRVSEKQRD
DVAMEKPYVEQTLASEACNTRNHVNPKNPKYGRSKDERRRRLSDVETETKRSY
GRKSDCSRMVNIPEKMEYKEESEESKLSNBEINKRVEEFIQRFNQIRSGSPRS
ST"

repeat_region

complement(8757..8780)
/rpt_family="AT_rich"
complement(10775..10811)
/rpt_family="AT_rich"
complement(10873..10905)
/rpt_family="AT_rich"
complement(11013..11070)
/rpt_family="AT_rich"
complement(11798..11818)
/rpt_family="AT_rich"
complement(13002..16330)
/gene="T2H7.12"

gene

/note="pseudogene, putative retroelement polypeptide"
/pseudo

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

mRNA

complement(115657..16345)
/rpt_family="ATCOP1A13|ATCOP1A13I Internal region of
ATCOP1A13 LTR-retrotransposon."
/rpt_family="ATCOP1A14|ATCOP1A14I Internal region of
ATCOP1A14 LTR-retrotransposon."
join(<17486..17518,17796..18535,18620..>18947)
/gene="T2H7.3"
17486..18947

gene

/note="predicted by genemark hmm"
join(17486..17518,17796..18535,18620..18947)
/gene="T2H7.3"
/codon_start=1

CDS

/product="hypothetical protein"
/protein_id="AAG50857.1"
/db_xref="GI:12321631"

/translation="MSRLSKLSPLIHKRSVRSFSSSTTGPIYLOLSAKRSSGGVYN
IGELIVLDPAPQELNLNDKPIPEIIVAKMIGASGWAFLIDQDRVLLITDLSNP
ACKSNPKLITLPLNPLFSQOTDIWVNAASSCDDEDMVVGKISLGDQVSEFRPR
DLRWKFKQTPSDHPTSNLTKSKDRKRYLPGGSHHLSYDLDPNKKADQEFHELOR
RNPEPSTKYDSELSLPSSSCRTSRFVSPSGDRPLVYKWAACGKIASSSKITYETOR
PMVREBETTERKACTYDDIGDLCIFYSKSEARCVPAASSTPGILKPNSTYVGFGLGI
YDLTRNVSTIFRAPKGAALQIVSPWPPASS"
complement(117528..17528)
/rpt_family="CAAAAAn"
complement(join(<19380..19710,19799..>20533))
/gene="T2H7.4"
complement(19380..20533)
/gene="T2H7.4"

repeat_region

mRNA

gene

/note="similar to hypothetical protein GI:8778720 from
(Arabidopsis thaliana)"

CDS
complement(join(19380..19710,19799..20403))
/gene="T2H7.4"
/codon_start=1
/product="unknown protein"
/protein_id="AAG50858.1"
/db_xref="GI:12321632"
/translation="MKLVYADKITYPEDELYDAQVYGAHQWGLFSDRTNRSAIISDYLN
PYASKSKPMHILPEFTPMYSQDEVCNVAMSPDPQDDHEDVMVGIKFLGROLS
LCRPHDLRWNTILTPESMEVSELYSKDORFLYLAQGYGLCSMDLNREDEKPK
FHELYHLNLPNPRSLMKOLDSYGRMDHVESPESEFLYMKFCEYSPENISNMCKSP
IYMPREBETKDGKRNMYTDDIDGLCFISGDEPCYKASSCPGLQNSLSHGRLP
AHLNLTKRTMGIEYFGIPIKRIPIPLWLPFSP
complement(21599..21620)
/rpt_family="AT-rich"
/complement(21745..21765)
/rpt_family="AT-rich"
complement(join(<21856..22030,22534..22806,23042..23208,
23290..23520,24442..>24837))
/gene="T2H7.5"
complement(21856..24837)
/gene="T2H7.5"
/note="similar to En Spm like transposon protein
GI:415361 from (Arabidopsis thaliana)"
complement(join(21856..22030,22534..22806,23042..23208,
23290..23520,24442..24837))
/gene="T2H7.5"
/codon_start=1
/product="En/Spm-like transposon protein, putative"
/db_xref="GI:12321633"
/translation="WSSSAIAAALKALERIRIMDVRSYVGLFSSVTVDNYSNRF
RKQCIQVIAKGAVALNIPDAVEVETLHNSMTCYIEIVHKTPMAAPKRGMPD

Query Match 41.7% Score 750.8; DB 8; Length 72058;
Best local Similarity 68.2% Pred. No. 5.2e-212;
Matches 1075; Conservative 0; Mismatches 492; Indels 9; Gaps 2;

QY 223 aagcccaagcaagaatccaaactcaacatacagatgaattcgtccagaagagcgcgcg 282
DB 55440 AACCCATATTCGCCAAGAACACTCTCCGTTGGAACCCCTCCACGCGCGCTCC 55381
QY 283 gcaagcttggaagcgcgagaggttcctcgttcagccaagagaagtaaccgctcct 342
DB 55380 GCCGCACTGCACTCGCCAAACCGCATTTGTAAGACGGCGATCTTAACTTCCTCC 55321
QY 343 aaaaagcgtatcctagctgttcagatcgcgcgaattcgtccgcgtgaatgaacagcc 402
DB 55320 AAAACAGTTGATCTCGCATGATCTCGGTAACACTAGCCTCGGTACCGAGCAATCC 55261
QY 403 gtccgagcgaatccttcgcgttggtcgaaaactccgattccaataaagagtgatgtg 462
DB 55260 GTTAAATCATCTCTCCCTGATGAGAAAATCCCTGACTGATGACGGTGTTCCTCC 55201
QY 463 cgaagagagcgttaaccacttcacgagccggtgacaggttcacactcttcgacgaagac 522
DB 55200 CGTAACGGCGCTAATCACTCTTCGAGCGATTTCTGTCATCACTATTTCGACGGTTC 55141
QY 523 ggtatggttcaagccgctcaaatcggaacaggttcacgactgaagcgttcgcggttact 582
DB 55140 GGTATGTTACAGCCCTTAAATCACTAAGAGAGACGCAAGTTACTCGTCCGGTTTAA 55081
QY 583 cagaataacggtttgttcagaagcgtcaattggtgcgacggtttccccaagaagcacc 642
DB 55080 GAAACCGAGAGATTGTTCAAGAGAAACAATCGGTTCTCCGATTTTCCCTTAAAGCTATA 55021
QY 643 ggttgagctcaagcgcaacacggatattgcggaactcaatgtatttcagccagagctgca 702
DB 55020 GGTGAGCTCAATGCTACTCTGGAATCGACAGGTTGATCTATTTCACACCTGGTTTAA 54961
QY 703 gccggtatagtcgacccgcaacggaacggtgtgactaaacgagcgttgtgtctattc 762
DB 54960 TTGCGTTTATTAAATCATCAAAAACGAGACCGGAGTTGCTAAGCCGGTTGGTTACTTCC 54901

QY 763 aatgagcgttatatggtcctatgctcagagatgattacccttaacgaattcagatccccc 822
DB 54900 CAGACCGGTTATTTAGCTATGTCTGAAGATGATCTACCTTACCAGTTGCTGCTCACTGAC 54841
QY 823 aatgagatcttaaaacggttggtcgttcgatttcgatttgatgacaattagatcccaatg 882
DB 54840 AATGGGATTTTATGAGACCTATCGAAGATTCGATTTTCAGAGGACCACTAAGCTCCCGCA 54781
QY 883 atgcccacccgaagtcgacccggaatccggttaactcttcgctttaagcagagctc 942
DB 54780 ATCGCTCACCCCAAGATTATTCGGTACGAAGAGACTATTTCGCTTGACCTACGACGTC 54721
QY 943 gttcaaaagccttaactaaataactccgattccacccggaacgaactaacacccgagc 1002
DB 54720 GTTAAAGAACCGTATTTGAATATCTTTAAATTTCTGCGCGAAGTGAGAAATCACCGGAC 54661
QY 1003 gtcgagatccagcttgatatacccaacagatgatagcagatttcgcatatcagaagactc 1062
DB 54660 GTTGAATTCCTCTCGCCACGTCGACGATGATGACGATTTCCGATTCACGATGAAATTC 54601
QY 1063 gtcgtcgtacccgagcaagatcgctttcgaagctgcgagagatgataccggtgtgtct 1122
DB 54600 GTTGTGATTCGGATCAACAAAGTTGTGTTAAGCTCTCGATATGTTTCTTGGAATCT 54541
QY 1123 ccggttggttcaagcaagaacaaggttcgaagattcgggatttgaacaatacgcgcaa 1182
DB 54540 CCGGTTAATATACGACGAGAGAAATTTCCCGTTTGAATTTTGCGCTTGAATACCGTAA 54481
QY 1183 gattatcgaaatcaatgaatgagatgtatgctccaagattgcttctgcttccattcggaa 1242
DB 54480 GATGCTCGCAATGTTGGTGTGATGATCACCGGAGACTTTTGTTCATCTTTGGAAAC 54421
QY 1243 gcttggaagagcagcaaaacagatgaatgcgtcgtatagatggttcctgtatgactcaca 1302
DB 54420 GCTTGGGATCAACCGAGACAGACAGAGGTTGATGATGATGATGATGATGATGATGATG 54361
QY 1303 gactcaatttcaacagatcgcagagaaactcaagagtgctcgtcgtcgtcgtcgtcgt 1362
DB 54360 GATTCGATCTTACGACGAGCGATGAGCTGACGATCAACAGTGTTCATGATGATGATG 54301
QY 1363 aatctcaaacccggtgaatcaactcgcgtccgctcgcgtatctccaagaagatcaaaagtc 1422
DB 54300 AATCTCAAAACCGGGAATCCACGACGACATATATTCCTCGGTGGTT--CAGATG 54244
QY 1423 aacctcgaagcagagatggttcaacagaaacatgctcgcgtcgtcgtcgtcgtcgttac 1482
DB 54243 AATCTGAAGCTGATATGTTAAACGGAATCTTCGCGGAGGAGAAACCGGTAACGATAC 54184
QY 1483 ctggtcttagcgcgagcgttggtcctaaagtcacagattcgtcgttaagttgactactact 1542
DB 54183 CTAGCCATAGCTGAACCGTGGCTTAAAGTCTCAGGATTCGTTAAAGTTGATCTTTCAC 54124
QY 1543 gtagaagttaaagaacatctttaagcggttaaccgtttaagggagagcctctgtttcc 1602
DB 54123 GGAAGATGAATAATCACTTTTACGGCGGTAAATAATGAGTGAACCTTTTCTTAA 54064
QY 1603 ccggt-----agaagagagagagagagagagatatacctcctgttcttcgacagac 1656
DB 54063 CCTAGAGGTTTGAATCTGACGAGAGATGACGGTTATATATATGCTGTTTTCACGAC 54004
QY 1657 gagaagacatgaaatcggagttacagatagttaacgcggttagacttagaggttgaagca 1716
DB 54003 GAGGAGAGTTGGGAATCGGAGCTTCAATATTGTTAAACGCCGTTACGTTGAACATAAGCA 53944
QY 1717 acggttaaacctccgtcaagaggttcgtacgattccaaggtacattcaatcgaatcgaatc 1776
DB 53943 ACCGTTAAATTCGCCGCAAGATACGATGATGTTGTTCCACGAGCACTTTCGTGAATTCG 53884
QY 1777 gatttgcggaagcag 1792
DB 53883 GATATGTTAAACCAAG 53868

[illegible]

QY	664	ggatctgcgcgaatcctatgctattctacgcagagctgcagccggtatagtgcgacgcga	723
Db	613	ggaaattccgacaccttttgatgctgtttttacgcacgtggcgctgtgtgctgtgattcaacaccaa	672
QY	724	cacggaacccggttgaagtaacccgggtttgctatttaagtcgcsygtattggtatag	783
Db	673	aaacggccgtggaaagtaacaaacccgggtttggtttacttttaataaacggcgttttagcttag	732
QY	784	tcgagagatgatttacccttaacaaagttccagatccatcccaatgagatllaanaacggt	843
Db	733	tcagaaagacgattttaccgttaccattttaaaattttacttcaaacccggcatctccaaacggt	792
QY	844	ggtcgtgtcgaattttgatbgaacaattagaatccacaatgatbgtcccaaccggaagtgcac	903
Db	793	ggacgctttacgattttcgaacggcacttttaaaattccgaattgatattacccgaaactggac	852
QY	904	ccggaatccggtgagacctcttcgcttaagtacgaagtcgtttccaagcttaaccttaaa	963
Db	853	ccggtttaccgaagagcgtttcaacgcttttagctacgacgtctgttaagaacctttacccgaaa	912
QY	964	tacttcgcgattccaccgagcaggaacaaataccacggagtcgagatttcgaattgacag	1022
Db	913	tactttcacattttctgcacacagacggcgttttaattccgcgggaattggacagattcccgctcgaaact	972
QY	1024	ccaacgagatgcacagatttcgcgattacagagaacttcgttcgttacctgtaaccagcaa	1083
Db	973	ccgaacgattttcaacgattttccgatataaacggagaaatttttgctgattttcgtatcaacaa	1033
QY	1084	gtcgttttcaagctgcgcgagagatgataccgcgtgtggtctcggttgttatacgaacaagac	1144
Db	1033	gtccgtttcaagcgtccgcggagatattttccggtaaaattctccggtttgttttcgcacggagaa	1093
QY	1144	aaggtccgaagaattccggagattttagaanaatacgcgaagaattcatcgaacatlaagtg	1204
Db	1093	aaggtttcccgattttggcgatattttcccaabagacgcgcacaaacattttctcagattattctgg	1153
QY	1204	attgatgtccagatttgcttcctgccttcacatctctggaacgcttgggagaagaccgaanaa	1264
Db	1153	gtgaacctctccggagacgcttctgttttcattctcggaattcattgggaattcattgggaattccggagacg	1213
QY	1264	gatgaagtcgttcgtgataaggttcctgtatagaactccacacgaactcaatttcaacgaagct	1324
Db	1213	gagcgagattttggggatccggtatccgttatgtccgcgcggcgaatttcaatttccaacgagaa	1273
QY	1324	gacgggaatctcaagagtgctctctgtgaataatccgccttgatattcaaaacggtgataca	1384
Db	1273	gacgagacatttgacaaagcgttttttctcggaagatcagagattaaacttcgaaacactttaaac	1333
QY	1384	actgcgcgtccgacatcattcccaagaagatcaacaagtcacacttcgaagcagagatgtgc	1444
Db	1333	acgcgtctgttcgtttgttggtttaacgagat-----gtaaattttacgatttggatatggtt	1388
QY	1444	aacagaacaatgctcgcgcgttaaaacaaatcgcttaacttgcttgactttagcgcgctgtg	1504
Db	1387	aaacggaaacccggtttaggaagaaacccgggttcgcgtttttggccttatgttatcctttgg	1444
QY	1504	ccctaaagctcgaagatcgcgttaaagttgatctcaactacttgagaaagttgaagaacgtc	1564
Db	1447	ccaaaagttttccggtttgccttaacgttgatcttttgcacccggtgaagatgaaaaaattattatt	1508
QY	1564	tacgcgcgtlaaacgcttlaacgagagagcctctcttctcccccgt-----agaa	1614
Db	1507	tacgcgcgttgagaaatatttgccggcgaaacccgtttttcttgcgcggcaaatccggttaacggc	1566
QY	1612	ggaagagagagaacgaagatactcctctgttctgcgttcaagaagagaacatgagaa	1672
Db	1567	gaagaaatgaagaatgacgcggtattattttttgttcacgcttattacacgaagaaacaaagaca	1628
QY	1672	tcgaggttaacgataagttaaacgcgtttaggttgaagttgaagcaacggttataactgcg	1732
Db	1627	tcagacgcttcagattatttttaacgcgttttaatttaatttaacgcttbaaccttccagatttaactacgc	1688
QY	1732	tcgaaggtctccgaattcgaatttcaacggtatactcaggaacgagatgatttggcgaaaga	1790

[illegible]

QY	364	cagatcgccggaaattlctgcctccgigaaatgaaacagcccgctccggcgtatcttcctg	423
Db	2300	CAGATTTGGCCGGGAATTTATTCCTCCGGTACCGGAATCTTCCGTCGGGGAAACCTCATCCGTC	2359
QY	424	gtcgaaacttcccgattccatccaagaggtgtatgtgcgaacgagactcaaccactt	483
Db	2360	GAAGGAACAATCCCTGACTCCTATGACGGGTTTATCTCCGTACAGCGCCGATCCGATG	2419
QY	484	cacgagccggtgaaagaatcaaccactctcttcgaacgaaacgatatggttcaacgcgtcaaa	543
Db	2420	TTTGAGCCCAACAGCTGGGACACATTTATTTGACGAGACGGAATGGTTTACGCATTTAA	2479
QY	544	ltcgaaacacggtltaagctaaagtaagcttgcggtltaactaaactaacccggtlgtcaag	603
Db	2480	ATAACCAACGGTTTCACTACTCTACGATGCGCGTTTACAAAAACCGAGATTTGGTTACG	2539
QY	604	gaacgtcaattgggttcgaacggtlctcccaagaacatcgtgtgagctltaacggtcaacc	663
Db	2540	GAAAAACGATTTGGTTCACCAAGTTTTCGCCAAAGCAATGGGAGCTTCACGCTACATCG	2599
QY	664	gtaattgcgcgaacacagcatctatctcaagcagaagctcgacgcggtatagltcgaccggtca	723
Db	2600	GGAAATCGACCTTTGATGTCGTGTTTACGACACGTGGGCTTTGGTGTCATCAACAACAA	2659
QY	724	cacggaacccggtltagctaaacgcggtlgtgtctatltcaatgagccggtltagctag	783
Db	2660	AACGCGCTCGGAGTAGCAAAACGCCGGTTTGGTTACTTAAATACCGGCTTTTACCTTAG	2719
QY	784	tcgagagatgattactcttccaaagltcgaatcaccccaatggagattttaaaccggt	843
Db	2720	TCAGAAAGCGATTTATCCGTCACCAATTAATAATTATCTCAAAACCGGCGATCTCCAAACCGTT	2779
QY	844	ggtcgtgttcgaatttttgaatgaaatcaatgaatcaatgatctgcccacccgaagctcac	903
Db	2780	GGAGCTTACGATTTTCAGACGGTCAGTTAAATCCGCAATCATATGCTACCCGAAACTGGAC	2839
QY	904	ccggaatccggttgaactcttcgctttaaagctacgacgctgcgtlccaagccttaacaa	963
Db	2840	CCGGTTTACGAAGGAGCTTTCACGCGTTTAACTCAGATCGATGCTTAAAGAAACCTTACTCGTAA	2899
QY	964	tacttcogattctcaacggagcgaactaaatacccggaacgttcgaattcgaattgataag	1023
Db	2900	TACTTCACATTTCTGCCACAGCGGGTTTAATCCCGGAATTTGAGATGCCGCTCCGAAACT	2959
QY	1024	ccaacgaatgacgacgaatttcggaattacagaagaactcgtcgtcgaactgaacagaac	1083
Db	2960	CCGAGACGATTCACGATTTCCGTATACGGAAGAAATTTGTGTGATTTCTGTATCAACAA	3019
QY	1084	gtcgtlctcaagctccggaagatgaltccggtgtggtgtcctcgtgtgtgtltaacgacaagac	1143
Db	3020	GTCGTTTCAAGCTCCGGCGAGATGATTTCCGTAATTTCCGGTTGTTTTCAGACGAGAA	3079
QY	1144	aaggttcgaagaattccggtatlttaagacaataacgcggaagattcatcgaaactaagtgtg	1203
Db	3080	AAGGTTTCCCATTTGGGGAATTAATGCCCAAGACGCACAGAAGCTTCTCAGTAATCTGG	3139
QY	1204	attgatgtccagaattgtctctcgtcttccatctcttgaaacgcttgaaagagacgagaacaa	1263
Db	3140	GTGAACCTCTCCGGAGAGCTTCTGTTTTCATCTCTGGAATGTCATGGGAATCGCCGGAGACG	3199
QY	1264	gatgaagtcgltcgtgataaggtctcgtatgactcaaccagactcaattltaacagactct	1323
Db	3200	GAGGATTTGGTGATTCGATCGATGCTGATGTCCCGCGGATTTCAATTTTCAACAGAGA	3259
QY	1324	gaacgaatctcaagagtgtcctcgtcttgaaatccgcctgaaatctcaaaacccggtatca	1383
Db	3260	GACGAGACTTGAACAACCTGTTTGTCTGCAATAGATTAACCTCGAACAACAGCTTAAACC	3319
QY	1384	actgcgcgtcgaatcatctccaacgaagatcaacaagtccaactcgaacgagcaggtgtct	1443
Db	3320	ACGCTCTTCTGTTGTTGTTTAAACGAGAT-----GTAAATTTAGACATTTGCTATGCTT	3373
QY	1444	aacgaagaacatgcctcgcgcgtlaaaaccaaattcgtctactgtgctttagccgaacgcgtg	1503

Db 3374 AACCGAACGGTTAGGAAGAAAACCCGGTTCGGTTTGGCTATTCCTTATCCTTGG 3433
Oy 1504 cctaaagctcagatcgcgtcctaaagtctcactactcgtgaggaagtaagaacatcct 1563
Db 3434 CCAAAAGTTTCGGTTCGCTAAGSTCGATCTTTGCACCGGAGATGAAAAATATATT 3493
Oy 1564 tacgcgataacgcgttaacgagagagcctctgttctcccg-----agaa 1611
Db 3494 TACGCGGTGAGAAATATAGCGGACCGGTTTCTTGCCCGCAACTCCGTAACGGC 3553
Oy 1612 ggaagagaggaagacgaagatacctcctcgttctcgttcaacgagagaacatgga 1671
Db 3554 GAAAGAAATGAGATGACGGTTATATTGTCACGTTCAAGACGAAGAAACAAAGACA 3613
Oy 1672 tcggagttacagatagtttaacgcgttagcgttagaggttaacgaacgttaacctcg 1731
Db 3614 TCAGAGCTTCAGATTATTTACCGCTGTATTTAAAGCTTGAAGCTTAACTACCG 3673
Oy 1732 tcaagggttcgcgtacgattcgcgtatcatcgcgagccgattgagcgaaga 1790
Db 3674 TCTAGAGTACCGTATGTTGCTTCATGCGACATTTGTGATTCGAATGCAACTCCTGATCA 3732

RESULT 15

LOCUS ATCHRIV48 194143 bp DNA linear PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 48.
ACCESSION AL161548
VERSION AL161548.2 GI:7268604
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 107966 to 194143)
Hilbert,H., Braun,M., Holzer,E., Brandt,A., Duesterhoeft,A.,
Mewes,H.W., Lemcke,K. and Mayer,K.F.X.

REFERENCE
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 38835; 28388 to 111084)
MURPHY,G., RIDLEY,P., HUDSON,S., MEWES,H.W., LEMCKE,K. and
MAYER,K.F.X.

REFERENCE
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 194143)
EU Arabidopsis sequencing, project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@brc.ac.uk

COMMENT
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
this fragment has an overlap with ATCHRIV47 at the 5' end and an
overlap with ATCHRIV49 at the 3' end.

FEATURES
SOURCE Location/Qualifiers

1. 194143
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="4"
6377..9293
/gene="AT4g18120"
join(6377..6796,7122..7238,7347..7501,7578..8456,
8543..8627,8716..8784,8871..9293)
/gene="AT4g18120"
/note="similarity to Me12-like protein, Arabidopsis
thaliana, gb:D86122
contains EST gb:W43784, A1998063.1"
/codon_start=1

/product="putative protein"
/protein_id="CAB78814.1"
/db_xref="GI:7268605"
/translation="MASGDHFGSGIGNMLDGDGSHPIGNMLPDDEELFSGLDLON
LSSLPARTDLEDYDLGSGGLFLEDDPYSLKMGSRMGFPASNDVNPOMITON
GVGSIAGHPYGEHPRSRLTYVRNTNSVDESELOALFEDNSEKDVNGTILVYNLAP
SVSNRDLNENIGVYGEIKERLETIPNKRHHKFEVEFYRSADALKALNREILGKRI
KLEHSRPGGARMMLOMNPLEDODDSYSLNHVESPLASSPIGNWRNSPIDHPLQSF
SKSP1FGNLSPTKNIRYDEFEMKTASVNDDEGRFESHLDLFSSSYNASHSKSTF
QOPOFSGSVSEGLNSHPSHVETLSGEFLMGSPSSAMVPNFSNREHREPYS
ONGSIHOLHHIGSAPSCGFPRSPETSSMGSYAFPGASGNMNAORNLRETSRPNKMLS
APRSQFLFTNGSYLMPRATVYSTDDPLEDSNQOFPDSNGQADIKIOPOLDKIMR
GEDPRTLMINKNIPNKYTRNMLLAIDKNSGTVDFTLYLPIDFNKCKVGAFTNWS
PKFTIAYEAFNGKRMKPFNSEKVASLARIQKALIAHFGNSSLMNEDRCPITV
FDGSESKYPVNNLLQCTHDLKMIYLFVCVYLMNMLFTHCLFETRSSEFFARASNTG
FLIFLSARTHIKADSQVNHVIAAGLKKRYQLG"
6377..6796
/gene="AT4g18120"
/number=1
6797..7121
/gene="AT4g18120"
/number=1
7122..7238
/gene="AT4g18120"
/number=2
7239..7346
/gene="AT4g18120"
/number=2
7347..7501
/gene="AT4g18120"
/number=3
7502..7577
/gene="AT4g18120"
/number=3
7578..8456
/gene="AT4g18120"
/number=4
8457..8542
/gene="AT4g18120"
/number=4
8543..8627
/gene="AT4g18120"
/number=5
8628..8715
/gene="AT4g18120"
/number=5
8716..8784
/gene="AT4g18120"
/number=6
8785..8870
/gene="AT4g18120"
/number=6
8871..9293
/gene="AT4g18120"
/number=7
complement(9594..9837)
/gene="AT4g18130"
/number=1
complement(join(9594..9837,9944..10237,10323..11130,
11238..13230))
/gene="AT4g18130"
9594..13230
/gene="AT4g18130"
complement(join(9594..9837,9944..10237,10323..11130,
11238..13230))
/gene="AT4g18130"
/note="Contains ATP/GTP-binding site motif A (P-loop)
AA634-641:Phytochrome chromophore attachment site
AA317-326
contains EST gb:r21020, AA395462"
/codon_start=1
/product="phytochrome E"
/protein_id="CAB78815.1"

`/db_xref="GI:7268606"`
`/translating="MGFESSSSSTAASNMKPQPKNSNTAQTAYSDAALFADFAGSITVTKRS`
`FNYSGVSPENHVPDEHITITVNIQGGGLVQPEFGCLIAVEEFPRLIGLSDMSDPF`
`LGILSIPSTSGSEDFKQKGLGIDARTLFPSSGASISKASFTSILNPLVPHSR`
`TQTKREYVILIRIDAGIYMDLEPAKSGDPALTLAGVAOSKALVRAISRLSILGCGD`
`GALCDTVEDYQRLTGYDRWYVOFHEDDHEVYSFIRSLDEYLIGHATPTIPQAR`
`ARLFQRKVRHRCIDCNATPVKVVQSEELKRPCLLVNSTLEAPHGCHTQYMANNGSVAA`
`SLIALIVKRGDSSKLMGLVGNHCSPPRYVPEPLRYACEFLMQAFGLQMLQELASQ`
`LAERKAMTQRLCLDMLLRDTVSAIVTQSPGIMDLVKCDGAAIYKKGCMVLGVTPMB`
`SOVKLVLMVLENNHGDSDGTGLTDSLDVAGVGAISLGDVAGVAAAEFSKDXILME`
`RSMTASAKKMGAGKHHPKDKDQDGMHPRSFTAFLEPAKSRSLPWEISSEIDAHSLR`
`LIMRESPTSSRPVLSGNCVARDANEMLTSFVCEMYRVLETATAPITFYDSSGTSIMNKN`
`KTAEKTTGLLASEAMGRSLADEIVQEESSPALESLLCKALQGEERKSVMLKRTKGONN`
`HPDVSIDCVSVNCSCTSDRYETENIGVCFQODITSEKAITDRFIRLOGDXYKTIVQSI`
`NPLDPIPAVDENACCSEMNAMAEKLTGMSHEYIKMLPGEVGFQCKYKQCSQILFR`
`PLISLYOGIADGNPESLIVPEFNKEGYIEASLTSKASTNIBEGVIRCEPFLIINKK`
`ESGSLPEKISASQINMETLYVROEIKRPLNGITPAHKLLESSTISASORFLETSDDA`
`CEKQJTTIESTDLKSIIEEKLQDETEERLENTLIDITISQVMTILLRNSQLNEVYA`
`EEIKTLPNGSRVAKLQILADLRLNIVHAFPPMSWGISISPEQELSRDGRVILHQL`
`FRMHHPGLSGEPMELSDMFEETRDGWPVDPDGLKLSRKLLEQMGNGRVSYVEDRECFEE`
`QVDLVQKTMILVESSEKTEGSSSSIK"`
`complement(9838, .9943)`
`/gene="AT4g18130"`
`/number=1`
`complement(9944, .10237)`
`/gene="AT4g18130"`
`/number=2`
`complement(10238, .10322)`
`/gene="AT4g18130"`
`/number=2`
`complement(10323, .11130)`
`/gene="AT4g18130"`
`/number=3`
`complement(11131, .11237)`
`/gene="AT4g18130"`
`/number=3`
`complement(11238, .13230)`
`/gene="AT4g18130"`
`/number=4`
`complement(14593, .14751)`
`/gene="AT4g18140"`
`/number=1`
`14593, .16312`
`/gene="AT4g18140"`
`complement(join(14593, .14751,14921, .15070,15338, .15639,`
`15915, .16055,16141, .16312))`
`/gene="AT4g18140"`
`complement(join(14593, .14751,14921, .15070,15338, .15639,`
`15915, .16055,16141, .16312))`
`/gene="AT4g18140"`
`/note="similarity to Hya22 protein, Homo sapiens,`
`PR2:JCS707`
`contains EST gb:R65235"`
`/codon_start=1`
`/product="putative protein"`
`/protein_id="CAB78816.1"`
`/db_xref="GI:7268607"`
`/translation="MKSRTKSDLSLEKRVIGVCKRPHKDVKNSCDPKTSRPMELIVY`
`ATQKNNNEQNIIGIDSSTRDYVYVNEPDLLVEDSSKASNMDTFSYVLDELXSET`
`DNEPVELLDAAEQIMLLPYLEDRSASDDVSDSDENRINRSRKMLEAENHTEAEQY`
`EDFQELVLRNPDELADVFNVPDMQOPRSPKRAVTFVLIWMLDILDPDGEFVQY`
`ORFYBDSQILSDGIYTKDLTVIGLDLAKVAIVDNCPOYRIQINGGIRIKSWYDPTT`
`DLILFLELETLDANDVPRVIAKRFNKKE"`
`complement(14752, .14920)`
`/gene="AT4g18140"`
`/number=1`
`complement(14921, .15070)`
`/gene="AT4g18140"`
`/number=2`
`complement(15071, .15337)`
`/gene="AT4g18140"`
`/number=2`
`complement(15338, .15639)`

[illegible]

```
Db 111045 AAGTTTCCCGATGCGGATTAATGCCAAGACGACAGAAAGCTTCAGATTAATCTGG 111104
QY 1204 attgatgtcccaagattgtctgtctccatctctgaaacgcttggaaagccagaaca 1263
Db 111105 GTGAACCTTCGCGAGCGTTCTTTCATCTCTGGATGCAATGGGATCGCCGGAGACG 111164
QY 1264 gatgaagtcgtcgtataagagtcgtatgactccaccagactcaattccaacgagctc 1323
Db 111165 GAGGAGATTGTGGTGTATCGGATCGTGTATGTCGCCGCGGATTCAATCTTCAACGAGAGA 111224
QY 1324 gacgagaatctcaagagtgctcctgtcgaatcccgctgaatctcaaaaacggtgaaatca 1383
Db 111225 GACGAGAGCTTAGAAGCGCTTTGTGCGAGATCAGGATTAACCTCAGAACACGTAAACCC 111284
QY 1384 actcgccgtccgactcccaagagaatcaacaagtcaacctgaaagcaggatgctc 1443
Db 111285 ACGCGTCGTTGTTGTTGGTTAACGAGAT-----GTAAATTTAGAGATTGGTATGTT 111338
QY 1444 aacagaaacatgtctcgccgtaaaacccaattcgcttacttgctttagcgcgagccgttg 1503
Db 111339 AACCGGAACCGGTTAGGAGAGAAAACCGGTTGCGGTTTGTGCTATTCCTTATCCTTGG 111398
QY 1504 cctaaagtcctcaggatctcgtctaaagltgactcactactcgaagaagttaagaacaatctt 1563
Db 111399 CCAAAAAGTTTCGCGTTCCGCTAAGTCGATCTTTGCACCGCGTGAGATGAAAAATATATT 111458
QY 1564 tacgagataaacgcttaaggagagagcctcgtctcctcccg-----agaa 1611
Db 111459 TACGGCGGTGAGAAATATGGCGGACCGCTTTTCTTCCCGCAACTCCGGTAACGGC 111518
QY 1612 ggaaggagaaagacgaagatacatcctctgttctgctacgaagagaacatgaaa 1671
Db 111519 GAAGAAAATGTGAAGATGACGTTATATTTGTGACGTTTCATGACGAGAAACAAGACA 111578
QY 1672 tcggaagttcaagatagttacgcccgttagcctaagaggttgaagcaacggttaacttcg 1731
Db 111579 TCAGAGCTTTCAGATTATTAACGCTGTATTTAAACCTTGAAGCTTACGATTAACCTACCG 111638
QY 1732 tcaaggttcgctacgagttcaagttacatcatcgcgaagcgaatgttgcgaaagca 1790
Db 111639 TCTAGAGTACCGTATGAGTTTCAATGACATTTTGATTCGATCACTCGTTGATCA 111697
```

Search completed: July 24, 2002, 07:42:25
Job time: 7835 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 04:30:44 ; Search time 224.85 Seconds
(without alignments)
13744.479 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800

Sequence: 1 atggtctcttcacgcgaac.....tggcgaacgacgtcgtgtga 1800

Scoring table:

IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1800	100.00	1800	22	AAD09396
2	824.6	45.8	1818	22	AAD09401
3	798.4	44.4	1839	22	AAD09399
4	738.2	41.0	1752	22	AAD09394
5	606.6	33.7	1815	22	AAD09400
6	528.2	29.3	1734	22	AAD09398
7	429.9	22.1	1734	22	AAC05678
8	204	11.3	492	22	AAC82706
9	188	10.4	372	21	AAC56595

10	171.8	9.5	325	21	AAC56548	Eucalyptus grandis
11	169.6	9.4	1788	22	AAC42989	Arabidopsis thaliana
12	169.6	9.4	1788	22	AAD09395	Arabidopsis thaliana
13	169.2	9.4	1950	22	AAE77206	CDNA encoding sunf
14	149.4	8.3	491	21	AAC57157	Pinus radiata tran
15	143.4	8.0	386	21	AAC57162	Pinus radiata tran
16	140.4	7.8	326	21	AAC57167	Pinus radiata tran
17	139.8	7.8	1617	22	AAD09397	Arabidopsis thaliana
18	139.8	7.8	1617	22	AAD09402	Arabidopsis thaliana
19	139.8	7.8	1777	22	AAC36083	Arabidopsis thaliana
20	137.2	7.6	329	21	AAC57165	Pinus radiata tran
21	121.6	6.8	398	21	AAC57145	Pinus radiata tran
22	118	6.6	340	21	AAC56520	Pinus radiata tran
23	113	6.3	238	21	AAC57169	Pinus radiata tran
24	109.6	6.1	246	21	AAC57168	Pinus radiata tran
25	106.6	5.9	412	21	AAC57164	Pinus radiata tran
26	89	4.9	190	21	AAC57163	Pinus radiata tran
27	78.6	4.4	303	21	AAC57125	Pinus radiata tran
28	76.8	4.3	308	21	AAC57142	Pinus radiata tran
29	74.8	4.2	395	21	AAC57117	Pinus radiata tran
30	67.8	3.8	285	21	AAC56908	Pinus radiata tran
31	65.8	3.7	567	21	AAC56546	Eucalyptus grandis
32	64.6	3.6	457	21	AAC56910	Pinus radiata tran
33	63	3.5	445	21	AAC56948	Pinus radiata tran
34	62.4	3.5	436	21	AAC57124	Pinus radiata tran
35	58.2	3.2	304	21	AAC56973	Pinus radiata tran
36	56.6	3.1	419	21	AAC56527	Eucalyptus grandis
37	56.6	3.1	447	21	AAC56480	Eucalyptus grandis
38	56.6	3.1	508	21	AAC56511	Eucalyptus grandis
39	55.2	3.1	384	21	AAC56540	Pinus radiata tran
40	53.8	3.0	311	21	AAC57151	Pinus radiata tran
41	49.4	2.7	321	21	AAC56510	Eucalyptus grandis
42	48.4	2.7	450	21	AAC56474	Eucalyptus grandis
43	46.4	2.6	501	21	AAC43547	Zea mays DNA fragm
44	45	2.5	384	21	AAC56556	Eucalyptus grandis
45	44	2.4	311	21	AAC56536	Eucalyptus grandis

ALIGNMENTS

RESULT 1	
AAD09396	
ID AAD09396	standard; cDNA; 1800 BP.
XX	
AC AAD09396;	
XX	
DT 10-SEP-2001	(first entry)
XX	
DE Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCE3 CDNA.	
KW Neoxanthin cleavage enzyme; ATNCE3; abscisic acid; ABA; herbicide;	
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
KW plant growth protectant; ss.	
XX	
OS Arabidopsis thaliana.	
XX	
EH Key	Location/Qualifiers
FT CDS	1..1800
FT	/tag= a
FT	/product= "Arabidopsis thaliana ATNCE3 protein"
XX	
PN EP116794-A2.	
XX	
PD 18-JUL-2001.	
XX	
PF 11-JAN-2001; 2001EP-0300218.	
XX	
PR 13-JAN-2000; 2000JP-0010056.	
PR 11-JAN-2001; 2001JP-0003476.	
XX	
PA (RIKE) RIKEN KK.	
XX	

PI Iuchi S, Kobayashi M, Shinozaki K;
XX
XX WPI: 2001-400081/43.
DR P-PSDB; AAE04784.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
producing transgenic plants with improved or decreased stress tolerance
PS
XX
XX
PS Claim 3: Page 32-36; 101pp; English.

CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE3 protein.
CC The AtNCE3 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPR65 (COWPEA Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.

XX
XX Sequence 1800 BP; 458 A; 464 C; 439 G; 439 T; 0 other;

Query Match 100.0%; Score 1800; DB 22; Length 1800;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1800; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctcttcacggaacgctgcgttctcggagatgcttggcgaatcact 60
DB 1 atggtctcttcacggaacgctgcgttctcggagatgcttggcgaatcact 60
QY 61 cagcgccattatcgtcttcacgaagctcgcacttgatgtatgtactccta 120
DB 61 cagcgccattatcgtcttcacgaagctcgcacttgatgtatgtactccta 120
QY 121 gccagtcgttcacgaagctcgaatgttcacgttcacacccctcact 180
DB 121 gccagtcgttcacgaagctcgaatgttcacgttcacacccctcact 180
QY 181 cattccctaagaatcacaactcccgcatgtgttaagcccaagcaagaa 240
DB 181 cattccctaagaatcacaactcccgcatgtgttaagcccaagcaagaa 240
QY 241 tccaaactaagaatgtattgttcacgaagcgcgcgcaagcttgagcg 300
DB 241 tccaaactaagaatgtattgttcacgaagcgcgcgcaagcttgagcg 300
QY 301 gaaggttctcttcacgacgagacgtacacccgcttctaacaagcggt 360
DB 301 gaaggttctcttcacgacgagacgtacacccgcttctaacaagcggt 360
QY 361 gtccaagtcgcggaatttgcgtcggatgaacacgctccggttaactc 420
DB 361 gtccaagtcgcggaatttgcgtcggatgaacacgctccggttaactc 420
QY 421 gtgttcggaataattcccggttcacaaagagtgatgtgcgcaacga 480
DB 421 gtgttcggaataattcccggttcacaaagagtgatgtgcgcaacga 480
QY 481 ctccaacgacggtgacagatcaccaactcttcgaacggaacggtatg 540
DB 481 ctccaacgacggtgacagatcaccaactcttcgaacggaacggtatg 540
QY 541 aaatcgaacacggttcagctacgcttgccggttactcagactaacgg 600

DB 541 aaatcgaacacggttcagctacgcttgccggttactcagactaacgg 600
QY 601 caggaacgtcaatttggctgcacccggtttccccaagacatcggtatc 660
DB 601 caggaacgtcaatttggctgcacccggtttccccaagacatcggtatc 660
QY 661 accggtatgcggaactcgaatcctaacgcagagctcgaacggatgac 720
DB 661 accggtatgcggaactcgaatcctaacgcagagctcgaacggatgac 720
QY 721 gcaacggaacggtgtagctaacgcggttggctatttcaatggcggtat 780
DB 721 gcaacggaacggtgtagctaacgcggttggctatttcaatggcggtat 780
QY 781 atctcggaagatatttactcctacaaagtacgaatcccaatgagat 840
DB 781 atctcggaagatatttactcctacaaagtacgaatcccaatgagat 840
QY 841 gtgttcggttcgaatttgaatgacaaatgaatcccaatgacccgaag 900
DB 841 gtgttcggttcgaatttgaatgacaaatgaatcccaatgacccgaag 900
QY 901 gaccggaacggtgtgaactcttcgcttaagctacgacggttcaaac 960
DB 901 gaccggaacggtgtgaactcttcgcttaagctacgacggttcaaac 960
QY 961 aaatcttcggaattcctcacaagcggaactaaatcccgagctgagat 1020
DB 961 aaatcttcggaattcctcacaagcggaactaaatcccgagctgagat 1020
QY 1021 cagccaacgatagtatgcgaattcgcgaatcagaagaacttcgtcgt 1080
DB 1021 cagccaacgatagtatgcgaattcgcgaatcagaagaacttcgtcgt 1080
QY 1081 caagtcgttccaaagctgcgcgagatgacgcggtgttcgttgcga 1140
DB 1081 caagtcgttccaaagctgcgcgagatgacgcggtgttcgttgcga 1140
QY 1141 aacaagtcgcaagatctggagatttgaagaataacggaagattcaga 1200
DB 1141 aacaagtcgcaagatctggagatttgaagaataacggaagattcaga 1200
QY 1201 tggatgtatgctcagaattctctcgttcacatctctggaacgct 1260
DB 1201 tggatgtatgctcagaattctctcgttcacatctctggaacgct 1260
QY 1261 acagatgaagtcgtcgtataggttcctgtatgaactcacaagact 1320
DB 1261 acagatgaagtcgtcgtataggttcctgtatgaactcacaagact 1320
QY 1321 tctgacgaagatctcagaatgtcctgtctggaatccgcgtgata 1380
DB 1321 tctgacgaagatctcagaatgtcctgtctggaatccgcgtgata 1380
QY 1381 tcaactcgcgttcgaatcactcacaagaagatcacaagctcgaag 1440
DB 1381 tcaactcgcgttcgaatcactcacaagaagatcacaagctcgaag 1440
QY 1441 gttcaacagaacatgctcggcgtaaaccaaatctcgcttactgt 1500
DB 1441 gttcaacagaacatgctcggcgtaaaccaaatctcgcttactgt 1500
QY 1501 tggccttaaatctcagaatctcgaatgtgatctactactggaaga 1560
DB 1501 tggccttaaatctcagaatctcgaatgtgatctactactggaaga 1560
QY 1561 ctttaacgcatataacggttaacgagagagccctgtttcccccgg 1620
DB 1561 ctttaacgcatataacggttaacgagagagccctgtttcccccgg 1620
QY 1621 gaagacgaagatatacatctctgttcttcgttcacgacgagaag 1680

Db 1621 gaagacgaagatatacctctctgttccttcacgacgagaacatggaatcgagttta 1680
Oy 1681 cagaatgaacgcgttagcttagaggttgaagcaacgcgttaacttcgtcaagggtt 1740
Db 1681 cagaatgaacgcgttagcttagaggttgaagcaacgcgttaacttcgtcaagggtt 1740
Oy 1741 ccgttagcgaattccacggttaacatccacgacgcgtatgttgcgaacgaagtcgttga 1800
Db 1741 ccgttagcgaattccacggttaacatccacgacgcgtatgttgcgaacgaagtcgttga 1800
RESULT 2
AAD09401
ID AAD09401 standard; cDNA; 1818 BP.
XX
AC AAD09401;
XX
DT 10-SEP-2001 (first entry)
XX
DE Lycopersicon esculentum neoxanthin cleavage enzyme, lenCED1 cDNA.
XX
KW Tomato; neoxanthin cleavage enzyme; lenCED1; abscisic acid; ABA;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; herbicide; ss.
XX
OS Lycopersicon esculentum.
XX
FH Key
FH Location/Qualifiers
FT 1..1818
FT CDS
FT /product= "Lycopersicon esculentum lenCED1 protein"
XX
PN EP116794-A2.
XX
PD 18-JUL-2001.
XX
PE 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI: 2001-400081/43.
DR P-PSDB; AAE04789.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
PT
XX
PS Claim 3; Page 67-71; 101pp; English.
XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An acid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Lycopersicon esculentum neoxanthin cleavage enzyme, lenCED1 protein
XX related to the invention.
XX
SQ Sequence 1818 BP; 569 A; 383 C; 369 G; 497 T; 0 other;

Query Match 45.8%; Score 824.6; DB 22; Length 1818;
Best Local Similarity 69.6%; Pred. No. 2,1e-256;
Matches 1134; Conservative 0; Mismatches 489; Indels 6; Gaps 1;
Oy 170 ctccagctcttcattcccttaagcaatcaatcaatccctccgcacattgttgaagcca 229
Db 185 ctccaattatcaaacaccaaagaataalacacattccaccccaaaaagaacaaca 244
Oy 230 aagcacaagaatccacaactaacaagatgaattgttccagaagcgcgcgacgct 289
Db 245 aatcctctctcttccactcccaagtggaatttagtgcaagacgacgacgaattgctt 304
Oy 290 tgaacgcgagcgaggtttctctcttcagccacgagaagctaacccggttccataaag 349
Db 305 tagatgtcgtagaagaagtgtcttaactaacaatgaacttgaaaccccttgcggaagaag 364
Oy 350 ctgactcagtggttcacagtcgcggaatattgtccggtgaatgaacacgcgtccgac 409
Db 365 ccgaaccacgagtcacagattcttggaatttgcgcgtacggaataatccagctgtc 424
Oy 410 gtaattccggtgttcggaatacttccgattccatcaagaagtgatgtgcgaacg 469
Db 425 aatctcttcggtacccggaataatcccaaatggttcaagcggttaccgttcgaacg 484
Oy 470 gagctaacccacttcaacgagcggtgacagtgacacacttcttcagcgagacgtatgg 529
Db 485 gagctaacccctcttcttgaaacacacgcgcgacacacattcttcagcgcgacggtatgg 544
Oy 530 ttacgcgcgtcaaatctgaacacggttcagctagctacgcttgcggttactcaagacta 589
Db 545 ttacgcgcgttcaattccaataatggtcgtcgtatgacgttgcggttccactgaacag 604
Oy 590 accggttgttcagaagaatcaattgggttcgacccggtttccccaagccatcgtaggc 649
Db 605 agaggtgttgcagaagaacgttgggtgcgcctgttcccttaaaagccatctgtgcat 664
Oy 650 ttcaagcgcacacggtatctccgcacatgctattcttaccagacgacgtcgaacggtta 709
Db 665 taactgtcactcttggaatttcgaagctatgctgtttaaagccgtggtcttcggac 724
Oy 710 taactgcacgcgcaacggaacccggtgtagtaacgcggttggctattcaattgagcc 769
Db 725 ttgtgtacacagtaagaagaactgtgttgcaaacgcggttagctattcacaataacc 784
Oy 770 ggtattgtcattgttcgagagatgatttaacttccaaagtcagatccatccaatgag 829
Db 785 gattactgtcattgttcgaagatgatttgccttaccatgtaagtaacacccacggtcg 844
Oy 830 attaaaaacgcgtgtcgttcgatttgaatgagcaattgaatccacaatgattggcc 889
Db 845 atcttaaacagaggtgtcgtattcgttcgacggtcagctaaatccacatgatagtc 904
Oy 890 accggaagtcgaaccgcggaatccggttgaactcttcgcttaagtaacgactcggttcaa 949
Db 905 acccaagctgcgaaccagattccggtgagctattgtccttagtaagatgtgattcaga 964
Oy 950 agccttacttaaatcttccgatttccacggaacggaactaaatccacggaactgaga 1009
Db 965 agccattccctaagtaattccacgatttcaaaaaaggggaaataacaaatgatatgttaa 1024
Oy 1010 ttacgctgacacgaacgagtgacagatttcgacgatttcagacgaagaacttcgtcgcg 1069
Db 1025 ttcaagtgaagaccacaatgacttgatgttccgacatttaccgagacacttcgtgcga 1084
Oy 1070 taactgacgaagctgttccaagctgcggaatgataccgcgtggtgtctccggttgg 1129
Db 1085 ttctcgtacacaagctgttccaagatgtcgaatgatccgtgaggtttcacgcgtgg 1144
Oy 1130 tttagacaagaagaagtcgaagatcggaattttagaanaatccgacgaagaatcat 1189
Db 1145 tttagacaagaagaagtttcccgatcttgatcttgataagaacggaagaatgtgt 1204
Oy 1190 cgaacattaagtgtatgtctccagattgcttctcgttccatctctggaacgcttgg 1249

[illegible]

XX	(RIKE) RIKEN KK.
PA	
XX	Iuchi S, Kobayashi M, Shinozaki K;
PI	
XX	
DR	WPI: 2001-400081/43.
XX	P-PSDB: AAE04787.
DR	
XX	A DNA encoding a protein with a neoxanthin cleavage activity for
PT	producing transgenic plants with improved or decreased stress tolerance
PR	-
XX	
PS	Claim 3; Page 53-56; 101pp; English.
CC	
XX	The invention relates to neoxanthin cleavage enzymes and their
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC	plant when expressed in a plant cell. The invention also relates to
CC	methods for increasing or decreasing stress tolerance in a plant by
CC	introducing the DNA into the plant, and a transgenic plant into which a
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid
CC	land can be improved by growing transformant weed for several years and
CC	then removing the weed by specifically lowering stress tolerance in the
CC	weed by inducing an inducible promoter. The present cDNA sequence encodes
CC	Vigna unguiculata neoxanthin cleavage enzyme, CPRD5 (Cowpea Responsive
CC	to Dehydration) protein. CPRD5 gene is isolated from cowpea plant.
XX	
SQ	Sequence 1839 BP; 447 A; 595 C; 442 G; 355 T; 0 other:
<hr/>	
Query Match	44.4%; Score 798.4; DB 22; Length 1839;
Best Local Similarity	70.4%; Pred. No. 6,7e+248;
Matches 1098; Conservative	0; Mismatches 456; Indels 6; Gaps 2
OY	233 ccaaaagatccacaacttaacagatgattgttccagagaagcggcggcagcttgg 292
Db	
	281 ccaaacaccattacctcaaaaatggaactttcttcagaaagcgcgtcacgccttgg 340
OY	293 acgcggcggaagggttctctctgtcagccacgagaagtacacccgcttctaagaagcgctg 352
Db	
	341 accttgctcgaacggcgctcgtctcgcacgaagcgaacaccgcgtccccaaaaagcgcg 400
OY	353 atccctagtttaagatcgcgcggaaatttgttcgccggtgatatbaaagcccgcgcgcgta 412
Db	
	401 acccggaagggtccaaaacgcgcgggaacttcgcgcgggttcggtagcatcgcgcgatcaa 460
OY	413 attcttcggttgcggaaaacttcccgatctccatccaagaagatgtatgtgcgcaacggag 472
Db	
	461 gactccgcgttgctcggaaaaatccccaatgatgatgacgycgtgtacgtgcgcgaacggtg 520
OY	473 ctAACccattcaCgaagccggtgacaggttcACaccacttcttcgaCggaagaCggtatgttc 532
Db	
	521 caatlcoggtctacagcctgtgcygcgcacacacttcttcgaCgCgagcgcatgttc 580
OY	533 acgcgcgtcaaatltcgaacCGgtltcaGctagctacgcttgcCGgtttactcagactaac 592
Db	
	581 acgcgcgtgaaglttcgaacCGgcgcgcgcaggtlacgcctgcCGcttcacCGgaCgagC 640
OY	593 gggttgtaagaagctcaatgggtgtgcagCGgtttccccaagccatCggttgaCttc 652
Db	
	641 gctcttcgacagagaaatctctcagcgcgcCGggtgttccCGaaggccatCggtgagcttc 700
OY	653 acggcacaacCGgtlctgtccgcatcatgtctatcttacgcgcgaagctgacgcggtatag 712
Db	
	701 acggcacaacCGgtcatcggcgctctctctcttcttcacgCgCgCGgtctcttcgggctcg 760
OY	713 tcgacCGgcgaCaCggaacCGgvtgtagtCaacCGcggttgggtctattctaatCgtgcggt 772
Db	
	761 ttgattggtgtccccaaggcatggcgctgtgcgaagcCGgtctcgtctactcaacaaCacc 820

QY	773	tattgctatgtcggagagatgatttaccataaagttcgaagttcgaatctccatcccaatgagatt	832
Db	821	ctctcggcccaagtcctccgaagacgattttaccctaccacacgtgagaatccacccctaaacgagcat	880
QY	833	taaaaaacggtttgtctggttcgcatcttgatgtgagacaattagaatctgaatccaaatgattgcccac	892
Db	881	taaacaccattgttcggtttacgacttccaaagcgagatctcaatccaaatgatctgcccacc	940
QY	893	cggaaagtcgagcccggaatccggtggaactcttcgctttaagctacgagctgcgtttcaaacg	952
Db	941	cgaaaactcggagcccccgtcgaacgagcctccacacgctccacagctacgacgtcatccaaagc	1000
QY	953	cttaacaaaataacttcgatatctccacccgagcgaactaaatcacccggagcgtcgaagttc	1011
Db	1001	cttaaccccaagttacttcgctttctcccccgaacggtcgaagttcccccgaacgtggaatcc	1066
QY	1013	agcttgatccagcccaacgatacgaatctcgagattcgagattacaagaaattctgcgtctac	1077
Db	1061	ccccgaaagagcccaacatgatgtacagattctgcgcataaagagaaatttcgtctgcctcc	1122
QY	1073	ctgacccagcaagtcgtttttaaagctcgcggagagatgtccgsggttggtctccggtgatt	1133
Db	1121	cggagccagcaggtggtcttccaacatacggagatgtacacacggcgggtgtcccccgtgctct	1188
QY	1133	acgacagaagaacaaagttcgcacaagatttcgggatttttaagacaatacgcgcgaagattatcga	1199
Db	1181	acgacagaagaacaaaactccacgtcttggtgtctctgcacaagaagtgcgaagacgcgaatg	1244
QY	1193	acataaagtgtgattgatgcctccagatgtgcttcctgcctccatctctcgtgaacgcttggaag	1255
Db	1241	cgaatgcggtggtatctcgaacgcgcgcgagatgtttcttcgcttccacctcttgaaacgcttggaag	1300
QY	1253	agcccaagaacagatgaaagctcgtcgtgaaggtccctcgtatgaattccacccaagatccaat	1311
Db	1301	agcccgaaacacggagaggttggtgtgattggtgcctcgaatagacccctcgtgagatccatt	1366
QY	1313	tcaacgagcttgacgagaaatccaaagatgctcgtctcgtatgaatccgcctgaaatccaa	1377
Db	1361	tcaacgagaaatcgaagagagattttgaagcgtctgctgtcgaagatataagcgtbaacttaagaa	1422
QY	1373	ccggtgatacaaatccgcgcgtccgcatactccaaacgaagaatacaacaaatcaaatccgaag	1433
Db	1421	ccggtcaagatccatccgcgcgtcccatctatctccgcgacgcg---aacaagtgaaacttggaag	1477
QY	1433	cagggtatggttcaacaagaacaatgctcgcgcgttaaaacccaatctgcgtacttggtttag	1499
Db	1478	ccggtcatgtgtgaacaagaacaagaatcgtgaagagaaacccagctcgcgtatctgtcgtcgtg	1533
QY	1493	ccggagccggtgcccctaaagtcttcagatattcgctaaagtgtgattccactactcgtggagaagtt	1555
Db	1538	cggagcccttcgcccacaagatctcggcgcttcgtcgaaagtgtgattcgtcgtgagtgaggaaagta	1599
QY	1553	agaaacatctcttaacgagcatcaaacccgttaccgagagagacccctcgtttctcccccgaagag	1611
Db	1598	agaagatcatgttatgtgaggaagagagaagttggtgtggtgagacccctcgtttcttccc---aaag	1655
QY	1613	gaggaagaggaagacgaaagatacatcctcgtttcgtttcgttccgaacgagaaagacatggaat	1677
Db	1655	gccaanaagaaagacgaatggtgtatattctgtgcattcgtgcgcgcgcgaagaaagaaatggaat	1711
QY	1673	cggaggttaacaatatgttaaagccggttgctctagaggtttgaagcgaagaggtttaaacttcgt	1733
Db	1715	ccggagctgcgaatcgtgtaaatgagcccaaaatttaagcgcgaagcttccatcaaaactccct	1777
QY	1733	caaaaggttcctcgtatggaatttcaacggtatcatctcggagacgcgaatgatttttgcgaaacgag	1799
Db	1775	ctcgtgttccctacagcttttcaatgatgaacttcatctcaatccaaagattttgagaaacaag	1833
RESULT 4			
AAD09394			
ID AAD09394 standard; cDNA; 1752 BP.			
XX			

AC	AAD09394,
XX	
DT	10-SEP-2001 (first entry)
XX	
DE	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 cDNA.
XX	
KW	Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;
KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW	plant growth protectant; ss.
XX	
OS	Arabidopsis thaliana.
XX	
FT	Key Location/Qualifiers
FT	CDS 1..1752
FT	/*tag= a
FT	/product= "Arabidopsis thaliana AtNCED1 protein"
XX	
PN	EP116794-A2.
PD	18-JUL-2001.
XX	
PF	11-JAN-2001; 2001EP-0300218.
XX	
PR	13-JAN-2000; 2000JP-0010056.
PR	11-JAN-2001; 2001JP-0003476.
XX	
PA	(RIKE) RIKEN KK.
XX	
PI	Iuchi S, Kobayashi M, Shinozaki K;
XX	
DR	WPI: 2001-400081/43.
DR	P-PSDB: AAE04782.
XX	
PT	A DNA encoding a protein with a neoxanthin cleavage activity for
PT	producing transgenic plants with improved or decreased stress tolerance
PT	-
XX	
PS	Claim 3; Page 18-22; 101pp; English.
XX	
CC	The invention relates to neoxanthin cleavage enzymes and their
CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC	plant when expressed in a plant cell. The invention also relates to
CC	methods for increasing or decreasing stress tolerance in a plant by
CC	introducing the DNA into the plant, and a transgenic plant into which a
CC	neoxanthin cleavage enzyme is introduced. The improvement of stress
CC	tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC	cleavage enzyme genes are useful for producing transgenic plants. An arid
CC	land can be improved by growing transformant weed for several years and
CC	then removing the weed by specifically lowering stress tolerance in the
CC	weed by inducing an inducible promoter. The present cDNA sequence encodes
CC	Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1 protein.
CC	The AtNCED1 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC	library using a cDNA of the CPRD5 (Cowpea Responsive to Dehydration)
CC	gene isolated from cowpea plant as a probe.
XX	
SQ	Sequence 1752 BP; 484 A; 396 C; 428 G; 444 T; 0 other;
XX	
Query Match	41.0%; Score 738.2; DB 22; Length 1752;
Best Local Similarity	68.2%; Pred. No. 2.2e-228;
Matches 1063; Conservative	0; Mismatches 478; Indels 18; Gaps 2;
OY	244 aaactaaacagatgaattgttccagaagcgcgccgcttctaataaacgcgtgatcttgatgt 303 Db 193aatcttcgccgtaaacactcttcgaagaagcgcgcgatcgatcgatcaagcggctgag 252
OY	304 gttctcttgtcaggccaagagaagctaacaccgcgcttctaataaacgcgtgatcttgatgt 363 Db 253 cgtgatatcatctcacagacaagaattctcaacttcccataaacgcgtgatccacagtgtt 312
OY	364 cgatccgcgggaatttttctccgltgaatgaagacccttcgcgcgtatattcttcoggt 423

Db	313	cagattccgaggaatttattcccggtaccggaattcttcgtccgacgaaacctcacgcgtc	372
Oy	424	gtcggaaaaacttcccgatttccattcaaaaggagtgatgtgacgaacgagatccacct	483
Db	373	gaaggaaacacccctgactgattgaagcggttatatactccgtaacggcggaatccgattg	432
Oy	484	caaggccggtgacagatgacacactctctcgacggagagcgttatgtgttcaacgcgtcaaa	543
Db	433	tttagcccaaacgcttggcaccaatttattcgcagagacggaatgtgttccagcagattaa	492
Oy	544	ttcgaaacggtttagcgttagttagcgtttgcggtttactcgaagctaacggtttgttaag	603
Db	493	ataaccaacggttcaagcttagctacgcatgcgcggttttcaaaaaacgcggagattgttcag	552
Oy	604	gaacgtcaattgggtgcgacgcggttttccccaagaacacgtgtagcttcaacggcacaac	663
Db	553	gaaaaacgattgggttcgacccagtttcccgaaagcaatgcggagcttccacggttcaactcg	612
Oy	664	ggtattgccgactacgtatgtattcttaacgccaagctgcaacgcggtatagtcgacccgca	723
Db	613	ggaactcgacgctttgattgctgttttaacgcacgctgggtcttggtctgtacacaacca	672
Oy	724	caacggaacccggtttagctaaacgcggttttggtctatttcaatgvcggttatgtgtatg	783
Db	673	aacgcgcgtcggatagatgacaaacgcggttgggtttacttacttaaacgcggtctttagctatg	732
Oy	784	tcggagagatgttttaaccttaacgaatttcaagatccacatccgatggagatttaaaacggt	843
Db	733	tcagaaagacggtttaaactcgtaaccaattaaattctcaaacgcggcgatctcccaaacggt	792
Oy	844	ggtcgggtcgaattttgatgatacgaattagaaatccacaatgtatgtcccaacccggaagtgcac	903
Db	793	ggaacttaggatttcgacgcgtcaggtttaaactccgcatgtatgactccacccaactgac	852
Oy	904	ccggaatccggttgaactcttcgctttaagtatacgtacgactggttccaagaacttaactaaa	963
Db	853	ccggttagaagaagacgttcaacgcggtttagctacgacgctgttaagaacaacttaactgaa	912
Oy	964	tacttcggaattccaacggacgagcaactcaaacacggaacgctgcagattgaactgtgttcag	102
Db	913	tacttcagattctcgcgaagacggttaaatcgcgcggaattggaagatcccgctccggaact	972
Oy	1024	ccaacagatgatacgaatttcgcgatatcagaagaaactcgtcgttaccctgacccaac	108
Db	973	ccgacgactgattcaacgaatttcgctataacggagaatttggtgtgattcctgatacaaa	103
Oy	1084	gtcgttttcaagctgcgagagatgataccgcggtgtgtctccggtgtgtttacgacaagaa	114
Db	1033	gtcgtgttccaagctcgcgagagatgatttccgcgttgaactctccggttgttttcgacaggaa	109
Oy	1144	aaggtcgaagaagatcggaatttgaacaaataacgcgaagaatcatccgaacattaaagtgg	120
Db	1093	aaggttcccgatctgggataatacgtcccaagacgcgacgaagaagcttccagataaactgg	115
Oy	1204	attgattgtccaagattgtcttcgttccatctctggaacgcttggagaagccagaaaca	126
Db	1153	gtgaactctccggagagcgtctgttttccatctctggaatgcatgaggatccgcggagacg	121
Oy	1264	gatgaagctcgtcgtgataaggttccctgtatgactccaacgaactcaattttcaacgagctc	132
Db	1213	gaagagatttggtgtatcgcgtatgtatgtcgcgcggcgagatcaattcaactcaacgagaga	127
Oy	1324	gacggaatctcgaagagtgctctcttgcgaatccgcgtgatactcaaaaacggtgatca	138
Db	1273	gacggagacttgagaagcggttttgvggagatccagataaacctccaagaacgtuaaac	133
Oy	1384	actcgcgttccgatactatccaagaagatataacaagtccaactcgaagacgagatggtc	144
Db	1333	acgcgtctgtctgttggtttaaagagat-----gtaatttagagattgtgtatggtc	138
Oy	1444	aacagaaacatgctcgcgcgttaaaacaaatctgacttactgtgctttagccgacgcgttg	150

Db	1387	aaccggagaacccggtgttagaagaagaaaaaacccggcttcgcgctttttgctcatgctgtatcctcttg	1446
Oy	1504	cctaagatctcagagatctgcctaagtctgatactcaactcgtgaagaagttaagaacaactct	1553
Db	1447	ccaaaagatcttcgcgcttttcgcgttaagatgcgtactctttgcaccggtgtgaagtaagaaataatact	1506
Oy	1564	tacggcgataacgctctacggagagagagccctctgtttctcccggtg-----aaga	1611
Db	1507	tacgctgcgtgagaatactatgctgcggcggaacgcttttctcttcgcccggcaactccgtaacgctc	1566
Oy	1612	ggagagagagagaagacggaagatacaatcctctcgtttctgttaccagacgagaagacatgtaaa	1671
Db	1567	gaagaaaaataaagatgacgctgtataatactttgtcagcttcatcgtacgtgaagaaacaaagaca	1636
Oy	1672	tccgagatcacaagatagcttaacgcttcagcttgaagctgtgaagcaacgctgttaactctgt	1731
Db	1677	tcaagcttcagatatacttaacgctgttaattttaaagcttgaagctcagatataaactacg	1686
Oy	1732	tcaaggtctccgtacggaatttcaacgctacatcatctcgaagccgcatgattgtgcgaagca	1790
Db	1687	tctagataccgctatgctgtttcttcattggtccacatctgtgtgactctgaagaactcgttatbata	1745

CC	RESULT	5	
CC	AAD09400		
CC	ID	AAD09400	standard; cDNA; 1815 BP.
CC	XX		
CC	AC	AAD09400;	
CC	XX		
CC	DT	10-SEP-2001	(first entry)
CC	XX		
CC	DE	Zea mays neoxanthin cleavage enzyme, VP14 cDNA.	
CC	XX		
CC	KW	Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;	
CC	KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;	
CC	KW	plant growth protectant; ss.	
CC	XX		
CC	OS	Zea mays.	
CC	XX		
CC	FH	Key	Location/Qualifiers
CC	FT	CDS	1..1815
CC	FT		/*tag= a
CC	FT		/product= "Zea mays VP14 protein"
CC	XX		
CC	PN	EP1116794-A2.	
CC	PD	18-JUL-2001.	
CC	XX		
CC	PF	11-JAN-2001; 2001EP-0300218.	
CC	XX		
CC	PR	13-JAN-2000; 2000JP-0010056.	
CC	PR	11-JAN-2001; 2001JP-0003476.	
CC	XX		
CC	PA	(RIKE) RIKEN KK.	
CC	PI	Iuchi S, Kobayashi M, Shinozaki K;	
CC	XX		
CC	DR	WPI: 2001-400081/43.	
CC	DR	P-PSDB: AAE04788.	
CC	XX		
CC	PT	A DNA encoding a protein with a neoxanthin cleavage activity for	
CC	PT	producing transgenic plants with improved or decreased stress tolerance	
CC	PT	-	
CC	XX		
CC	PS	Claim 3; Page 60-64; 101pp; English.	
CC	XX		
CC	CC	The invention relates to neoxanthin cleavage enzymes and their	
CC	CC	corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key	
CC	CC	role in endogenous abscisic acid (ABA) biosynthesis under drought stress.	
CC	CC	Neoxanthin cleavage enzyme is used for improving stress tolerance in a	
CC	CC	plant when expressed in a plant cell. The invention also relates to	
CC	CC	methods for increasing or decreasing stress tolerance in a plant by	
CC	CC	introducing the DNA into the plant, and a transgenic plant into which a	

neoxanthin cleavage enzyme is introduced. The improvement of stress tolerance in plants is useful, for example in plant breeding. Neoxanthin cleavage enzyme genes are useful for producing transgenic plants. An arid land can be improved by growing transformant weed for several years and then removing the weed by specifically lowering stress tolerance in the weed by inducing an inducible promoter. The present cDNA sequence encodes zea mays neoxanthin cleavage enzyme, Vp14 protein related to the invention.

Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 other;

Query Match	33.7%	Score 606.6	DB 22	Length 1815
Best Local Similarity	63.5%	Pred. No. 1.1e-185		
Matches 1003	0	Mismatches 549	Indels 27	Gaps 4

[illegible][illegible]

RESULT	6
AAD09398	
ID	AAD09398 standard; cDNA; 1734 BP.

AA AAD09398;
AC

DT 10-SEP-2001 (first entry)

Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED5 cDNA.

KW Neoxanthin cleavage enzyme; AtNCE5; abscisic acid; ABA; herbicide;

KW plant growth protectant; ss.

05 *Arabidopsis thaliana*.

Key	Location/Qualifiers
FH	

000	1:1:1:0:1	a
FT	/*tag=	a

XX

Query Match 11.3%; Score 204; DB 22; Length 492;
 Best Local Similarity 69.7%; Pred. No. 2.1e-55;
 Matches 304; Conservative 0; Mismatches 130; Indels 2; Gaps 2;

```

OY 627 ttccccaagccatcggtgagcttcaacgacacac-ggatltgcccgaactcatat 685
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 435 TTTTCTTAAGGAGATGAGTGCATGCGACATCGCGGAGTCCGCCCTTGCTCTGT 376
OY 666 tctaagccgaagctgagcggatagtgagcccggaacggagacgggtgagctaaag 745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 375 TCTACGCGCGCGCGCCG-CTCGGCTCTCTTGACCGGACACGCGACCGCGCTCGCCACAG 317
OY 746 ccgggttggcttattcaatgagccgttatgtgctatgctgagagatgattacattacc 805
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 CCGGCTCTATCTACTTCAACGCGAGGCTCTCGCATGTGCGAGAGACGACTCCCTTACC 257
OY 806 aagttcaagatcaactcccaatgagatttaaaaaacggtgtgctggttgaatttgatgagac 865
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 AGGTGCGCTACACCGCGGACGCGACTCGAGACCGTGGCGGCTACGACTTGCAGCGGCG 197
OY 866 aattggaatcccaatgatgtgcccaacccgaaagtgcgaccggaatccggtgaactcttg 925
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 ACCTGCGGCTGCGCCATGATCGCGACCCCAAGCTCGACCCGCGGACCGAGAGCTCCACAG 137
OY 926 cttaagctgaagctgctgttcaaaagccttaacataacttaccgattctcaacggagag 985
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 136 CGCTCAGCTACGACGATGATCAAGAACCGCTACCTCAAGTACTTCTTCTGCGGCCGAGAG 77
OY 966 gaactaaatccacggaagcgtgcgagatcagcttgatcgaacgaagatgagcagatttg 1045
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 76 GCACCAAGTCGCGCGCGAGTCGAGATCCGCTCGACAGCCACACATGATCAAGACTTCC 17
OY 1046 cgattacagagaactt 1061
    ||| ||| ||| ||| |||
DB 16 GAATTAAGTGAATCTT 1

```

RESULT 9

AAC56695
 ID AAC56695 standard; DNA; 372 BP.

AC AAC56695;

DT 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor DNA sequence #566.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; SS.

XX Eucalyptus grandis.

XX WO200053724-A2.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06112.

XX 11-MAR-1999; 99US-0266513.

XX 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI; 2000-579369/54.

PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine;
 PT having modified gene expression or modified activity of a polypeptide
 PT -
 XX
 PS Claim 1; Page 494; 747pp; English.

XX The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

CC Sequence 372 BP; 60 A; 148 C; 101 G; 63 T; 0 other;

Query Match 10.4%; Score 188; DB 21; Length 372;
 Best Local Similarity 69.1%; Pred. No. 2.8e-50;
 Matches 257; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

```

OY 639 catcggtagcttcaagccacacccggtattgcccgaactcatcttcaacgacagagc 698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 catcggtagcttcaagccacacccggtattgcccgaactcatcttcaacgacagagc 60
OY 699 tgcagcggatagtgagcccggaacggaacgggtgagtaacgacgggttggtcta 758
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 cctcttcggccttcgtgcacacccgaatgcatggtgcgtcgcgaagacggcgtctgta 120
OY 759 ttcaatgacgggtattggtcatgtcggagatgattacattaccgaattcagatcac 818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 ctctgagcgccactctcgtcgtcgcgagacgacccctaccagctgcgtcac 180
OY 819 tcccaatgagatttaaaaacggtgtgctggttcgatttgatgagacaattgaattcac 878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 ggcctccgggaactcgaagacgctgcgcgactacggttcgcgcgaactcgaattccc 240
OY 879 aatgattgccaccccgaaagctgacccggaatccggtgacatttcgcttaagctaga 938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 gatgatgcgccacccgaagaatgacccggttcgcgcgagatgttcgcctcgaactaga 300
OY 939 cgtcgtttcaagccttaccataaatacttcogattctcaacgagacgaactaatacc 998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 cgtcgtccggaagcgttaccataaatacttcogattctcaacgagacgaactaatacc 360
OY 999 ggaagtcagat 1010
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 cgaagtcagat 372

```

RESULT 10

AAC56548
 ID AAC56548 standard; DNA; 325 BP.

AC AAC56548;

DT 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor DNA sequence #419.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; SS.

XX Eucalyptus grandis.

PR	18-AUG-1999;	99005-01339465;
PR	18-AUG-1999;	99005-01337560;
PR	18-AUG-1999;	99005-01337673;
PR	21-AUG-1999;	99005-01338817;
PR	21-AUG-1999;	99005-01338819;
PR	23-AUG-1999;	99005-01403354;
PR	23-AUG-1999;	99005-01403355;
PR	24-AUG-1999;	99005-01406695;
PR	28-AUG-1999;	99005-01408823;
PR	29-AUG-1999;	99005-01409091;
PR	30-AUG-1999;	99005-01412827;
PR	01-AUG-1999;	99005-01418482;
PR	02-AUG-1999;	99005-01421554;
PR	02-AUG-1999;	99005-01420555;
PR	06-AUG-1999;	99005-01423390;
PR	08-AUG-1999;	99005-01428033;
PR	09-AUG-1999;	99005-01429200;
PR	12-AUG-1999;	99005-01439777;
PR	13-AUG-1999;	99005-01435422;
PR	14-AUG-1999;	99005-01435624;
PR	15-AUG-1999;	99005-01440005;
PR	16-AUG-1999;	99005-01440855;
PR	16-AUG-1999;	99005-01444085;
PR	19-AUG-1999;	99005-01443332;
PR	19-AUG-1999;	99005-01443332;
PR	19-AUG-1999;	99005-01443332;
PR	19-AUG-1999;	99005-01443332;
PR	19-AUG-1999;	99005-01443334;
PR	19-AUG-1999;	99005-01443335;
PR	20-AUG-1999;	99005-01443352;
PR	20-AUG-1999;	99005-01446322;
PR	20-AUG-1999;	99005-01446884;
PR	21-AUG-1999;	99005-01448814;
PR	21-AUG-1999;	99005-01448816;
PR	21-AUG-1999;	99005-01448816;
PR	21-AUG-1999;	99005-01450885;
PR	22-AUG-1999;	99005-01450885;
PR	22-AUG-1999;	99005-01450877;
PR	22-AUG-1999;	99005-01450889;
PR	22-AUG-1999;	99005-01451082;
PR	23-AUG-1999;	99005-01451922;
PR	23-AUG-1999;	99005-01454515;
PR	23-AUG-1999;	99005-01454516;
PR	23-AUG-1999;	99005-01454524;
PR	26-AUG-1999;	99005-01452763;
PR	27-AUG-1999;	99005-01453276;
PR	27-AUG-1999;	99005-01453913;
PR	27-AUG-1999;	99005-01459518;
PR	28-AUG-1999;	99005-01459518;
PR	02-AUG-1999;	99005-01453886;
PR	02-AUG-1999;	99005-01453886;
PR	02-AUG-1999;	99005-01453888;
PR	03-AUG-1999;	99005-01453889;
PR	03-AUG-1999;	99005-01453889;
PR	04-AUG-1999;	99005-01472024;
PR	04-AUG-1999;	99005-01473202;
PR	05-AUG-1999;	99005-01471922;
PR	05-AUG-1999;	99005-01472600;
PR	06-AUG-1999;	99005-01473063;
PR	06-AUG-1999;	99005-01474146;
PR	06-AUG-1999;	99005-01474149;
PR	09-AUG-1999;	99005-01479355;
PR	10-AUG-1999;	99005-01481771;
PR	11-AUG-1999;	99005-01483119;
PR	12-AUG-1999;	99005-01483341;
PR	13-AUG-1999;	99005-01485555;
PR	13-AUG-1999;	99005-01486884;
PR	16-AUG-1999;	99005-01493368;
PR	17-AUG-1999;	99005-01493175;
PR	18-AUG-1999;	99005-01494266;
PR	20-AUG-1999;	99005-01497222;
PR	20-AUG-1999;	99005-01497223;
PR	20-AUG-1999;	99005-01498299;
PR	23-AUG-1999;	99005-01499022;
PR	25-AUG-1999;	99005-01499300;
PR	25-AUG-1999;	99005-01505666;

PR	26-AUG-1999;	9905-0150884;
PR	27-AUG-1999;	9905-0151065;
PR	27-AUG-1999;	9905-0151066;
PR	27-AUG-1999;	9905-0151080;
PR	30-AUG-1999;	9905-0151083;
PR	31-AUG-1999;	9905-0151438;
PR	01-SEP-1999;	9905-0151399;
PR	07-SEP-1999;	9905-0152363;
PR	10-SEP-1999;	9905-0153070;
PR	13-SEP-1999;	9905-0153758;
PR	15-SEP-1999;	9905-0154018;
PR	16-SEP-1999;	9905-0154039;
PR	20-SEP-1999;	9905-0154779;
PR	22-SEP-1999;	9905-0155139;
PR	23-SEP-1999;	9905-0155486;
PR	24-SEP-1999;	9905-0155659;
PR	28-SEP-1999;	9905-0156458;
PR	29-SEP-1999;	9905-0156596;
PR	04-OCT-1999;	9905-0157117;
PR	05-OCT-1999;	9905-0157573;
PR	06-OCT-1999;	9905-0158029;
PR	07-OCT-1999;	9905-0158232;
PR	08-OCT-1999;	9905-0158369;
PR	12-OCT-1999;	9905-01585293;
PR	13-OCT-1999;	9905-0159294;
PR	13-OCT-1999;	9905-0159325;
PR	14-OCT-1999;	9905-0159329;
PR	14-OCT-1999;	9905-0159331;
PR	14-OCT-1999;	9905-0159637;
PR	14-OCT-1999;	9905-0159638;
PR	18-OCT-1999;	9905-0159584;
PR	21-OCT-1999;	9905-0160741;
PR	21-OCT-1999;	9905-0160767;
PR	21-OCT-1999;	9905-0160768;
PR	21-OCT-1999;	9905-0160770;
PR	21-OCT-1999;	9905-0160815;
PR	22-OCT-1999;	9905-0160980;
PR	22-OCT-1999;	9905-0160981;
PR	22-OCT-1999;	9905-0160981;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161405;
PR	25-OCT-1999;	9905-0161406;
PR	26-OCT-1999;	9905-0161359;
PR	26-OCT-1999;	9905-0161360;
PR	26-OCT-1999;	9905-0161361;
PR	28-OCT-1999;	9905-0161920;
PR	28-OCT-1999;	9905-0161922;
PR	29-OCT-1999;	9905-0161993;
PR	29-OCT-1999;	9905-0162142;

Query Match	9.4%;	Score 169.6;	DB 21;	Length 1788;
Best Local Similarity	50.2%;	Pred. No. 6.5e-44;		
Matches 709;	Conservative 0;	Mismatches 634;	Indels 69;	Gaps 9;

QY	426	cggaaacttcgcgattccatcaataagagatgatgtgcgaacgagagtaaccaactta	4855
Db	393	cggcacttcctacactgtacttaacgagcgcttaataccgtatccgtacgtccaatccacggtt	4522
QY	486	cgaagcgggtgacaggttcaacactcttcgaagagagacgtatgtttacgcggtcaat	5455
Db	453	tctccctcgtggtcccttaacacatccttcgaagcgagcgtatgcttcaagccaataaat	5122
QY	546	cgaacacggttcacgtatgactaagctctgcggttttaccgaactaacacggtttgttcaaga	6050
Db	513	ccaacacggttaaaagcaccatctctctgtagcagatgaagctataataataaacgycga	5722
QY	606	acgttaattgttgcgaccggttttccccaagaacatcgtgtgacttcaagcgccaacgcg-	6644
Db	573	gaacaacacgagagctccggttatgactcaaacggttttcccgatcaacaggttgaaagcg	6322

```
QY 665 --gatttcccgactcatgtctattcttaacgcagagctgcagccgtatagtcacccgac 722
Db 633 gtacagtagctctgtgaggtcttaacgcagcgtctagggttcttaacgcgacagtataatccggt 692
QY 723 aacaggaacggtgttagcttaacgcggtttgtctattcttaacgtgacggtatttgcgtat 782
Db 693 taacgcgcatgtgttagctataacaaagcttagcttcttcagtaacgcgtctcttgcgtt 752
QY 783 gtccgaggaatatttaccatttccaaagttcacatctcccaatgtagattttaaaccgt 842
Db 753 aggtgaatctgtattaccctacgcgtccgtacgttaacccaatcagagatttgaagacgt 812
QY 843 tgtctgcttcgatttcttgatggaacattagaaatcccaatgattgtcccccgaagctca 902
Db 813 cggacgcgtacgattctgcgcggaattagtcgattgatacagctatctccaaaccga 872
QY 903 ccggaatccggtgaactcttcgctttagtcgaagctgcgtttcaaaaccttaccataa 962
Db 873 tccaataacccggaagaaatttcgcttccggtacggtccggttcca---ccgttttaac 929
QY 963 atacttcgatttccacgcggaacgaactaaatccagcagctcgaga---ttcacctga 1019
Db 930 atattccggttgaattccgcgcggaataaaacaaagacggtccgataattctcgatgac 989
QY 1020 tcaagccaacgattgacgcagatttcgcgattacagagaaacttcgctgcgtacccga 1079
Db 990 gtctccgcgtcttccatcagcttcgcgatacgaacgtaacgagatttccgacagat 1049
QY 1080 gcaagctg-----ttccaagctgcggaagatgataccggtgtggtccggtcgt 1130
Db 1050 tcaagcttgcataggaagacatgttgattgttctcgaaggtgtgtcccggtcgt 1109
QY 1131 ttacgacaagaagacggaagattcgggaattttagaacaattccgcgcgaagatcatc 1190
Db 1110 tactgataacggaataaactcaaggtctgagtgatctcctaagtaacgcggaagacgagtc 1169
QY 1191 gaacattagtgatgattgattgctccagattgtcttcgtccatcccttgaacgcttgga 1250
Db 1170 ggagatgaagatggttcgaggtcccggtatcaataatcaatccagatataatgcttgga 1229
QY 1251 agagccagaacagatgaaagtcgctgtatgaggttcctgtatgactccaacagctcaat 1310
Db 1230 tgaagatgattgaaacagcgtcttcttgatgcacggaatattatgcatgataacatac 1289
QY 1311 ttccaagagctgaagagatctcaagagtgctcgtctgaacccgcgcgaactca 1370
Db 1290 tttagagagagatgatactcgtctc---atgcttctgtgtgagaaggtcgaatcgtcgt 1346
QY 1371 aaccggtgaatacactgcgcgtccgatacatctccaacgaagatacaacagtcacactca 1430
Db 1347 caccggtattgtgagaagctacgtactcagtcagcagag-----aatctcga 1391
QY 1431 agcaggagatggtcaacagaacatgctcgcgcgttaaacccaattcgttacttgcgtt 1490
Db 1392 ttccgctgtgatttaactccgaggttctcgtggaagatgatacgcggttcttaccgcgcat 1451
QY 1491 agccgcgcggtgcttaagctcagagattcgtctaaagttatctcactacgtcggaaggt 1550
Db 1452 tggagatcgcgttcgcgaagatctccggtcgtgtgtgaagcttgaatggtctcaaaagagacg 1511
QY 1551 taagaatacat-----cttaccgagataacggttaccggaagagacccctc 1595
Db 1512 ggaatgattgacgtgtgcccgttagaattgacggttcaggttgaacgcggaagacggtt 1571
QY 1596 gttcttcccgagaga-----aggaagagaggaagacgaagacatacatcctctgt 1643
Db 1572 ttctgtaactgaggaatcctgttaacccggaagcggaagagagatgattgttattgtgtac 1631
QY 1644 ttctgtaacgaggaagacatggaatcgaggttacagatagattgaacccggttagctt 1703
Db 1632 gtatgttcaacgataagagatggaagatcgaaagttctctgtgtgacgacgcaataatcgc 1691
QY 1704 agaaggttga-----gcaacggttaaaacttcgcgtcaagaggttccgtacgattcacg 1757
```

```
Db 1692 ggaagcttgaataacgtcgcgcgcgttgaggttgcgcggaaggttccgtacgattccatg 1751
```

```
QY 1758 tactatcatcggaaccgatgatttgcgaagc 1789
Db 1752 gtatttgcgaaggaagtgaccttaataagc 1783
```

RESULT 12

ABAD09395 standard: cDNA; 1788 BP.

AC ABAD09395;

DT 10-SEP-2001 (first entry)

DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE2 cDNA.

KW Neoxanthin cleavage enzyme; AtNCE2; abscisic acid; ABA; herbicide;

KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;

KW plant growth protectant; ss.

OS Arabidopsis thaliana.

FN Key Location/Qualifiers

FT CDS 1..1788

FT /tag= a

FT /product= "Arabidopsis thaliana AtNCE2 protein"

PD 18-JUL-2001.

PF 11-JAN-2001; 2001EP-0300218.

PR 13-JAN-2000; 2000JP-0010056.

PR 11-JAN-2001; 2001JP-0003476.

PA (RIKE) RIKEN KK.

PI Iuchi S, Kobayashi M, Shinozaki K;

DR WPI; 2001-400081/43.

DR P-PSDB; AAE04783.

PT A DNA encoding a protein with a neoxanthin cleavage activity for

PT producing transgenic plants with improved or decreased stress tolerance

PS Example 10; Page 25-29; 101pp; English.

CC The invention relates to neoxanthin cleavage enzymes and their

CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key

CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.

CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a

CC plant when expressed in a plant cell. The invention also relates to

CC methods for increasing or decreasing stress tolerance in a plant by

CC introducing the DNA into the plant, and a transgenic plant into which a

CC neoxanthin cleavage enzyme is introduced. The improvement of stress

CC tolerance in plants is useful, for example in plant breeding. Neoxanthin

CC cleavage enzyme genes are useful for producing transgenic plants. An arid

CC land can be improved by growing transformant weed for several years and

CC then removing the weed by specifically lowering stress tolerance in the

CC weed by inducing an inducible promoter. The present cDNA sequence encodes

CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCE2 protein.

CC The AtNCE2 cDNA is obtained from an Arabidopsis plant-derived cDNA

CC library using a cDNA of the CPD65 (Cowpea Responsive to Dehydration)

CC gene isolated from cowpea plant as a probe.

SS Sequence 1788 BP; 444 A; 444 C; 418 G; 482 T; 0 other;

Query Match

9.4%; Score 169.6; DB 22; Length 1788;

Best Local Similarity 50.2%; Pred. No. 6.5e-44;
Matches 709; Conservative 0; Mismatches 634; Indels 69; Gaps 9;

```

OY 426 cggaaactccgactcatcaaaagatgtatgtcgcaacggagcgaaccacttca 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 cggacactctcacgtctcaacggccttaaccgttaacgcgttaacccaatccacagtt 452
OY 486 cgaacggtgaaggtcacactctctcgaacgagatgtatgttcaacgcgttcaatt 545
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 tctccctgtgtctctaccatctctcgacggcgacggttatgtctcaacataaatt 512
OY 546 cgaacaggttaagtagtaagctgtgcggtttactcaagactaacgggttcttgaaga 605
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 513 ccaacacggtlaaagcaccctctgttagcagatacgaagactataataacaacgtcga 572
OY 606 acgtcaattgggtcgacgcggttttccccaagccatcgtgtgcttcaacggcacaacgg- 664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 gaacaaacacgagctccggttatacgtcttaacggtttccggtttaacggttgaacggc 632
OY 665 -glatgcccacatcatctatcttacgcagagctgcagccggtatagtcgacccggc 722
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 633 gtcagtagctctgtgaagctttaaaggcagctaggttttaacggcagatataatccggt 692
OY 723 aacaggaaacggtgtatagtaacgcggttctgtctattcaatggccggtatttgctat 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 693 taacgcatgtgttagtaataacaagctcagcttctctcagtaaacgctctcttgctt 752
OY 783 gtcggaaggtatttacccttaacgaattagatcaccatcgaatgtgaattaaacggt 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 753 aggtgaattcgtatttaaccctacgcgcgcgataacccaacgaatcgaatgaattgaacgat 812
OY 843 tgtctcgttcgatttttgatggaacaattagaatccacaatgatgtgccaccgcgaagtcga 902
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 813 cggacggtatcaggttctgaacgggaatattagatgattgaacgctcatctctaaacgga 872
OY 903 ccggaatccggttgaactcttcgctttaaagctacgaagctgttcaaacgcttaactaa 962
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 873 tccaataacggaggaactcttcgcttcggtacggtccggttcca---cgtttttaac 929
OY 963 atactctcattctcaccggaacctaataccaacggcgctcagaga---ttcaagttga 1019
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 930 atatttcggtttagtccgcggggaaaaaaacaagaagcgttcgatatcttcgattgc 989
OY 1020 tcaagcaacgatgatgcagatttcgcgataacagaacattcgtcgtacatgaacca 1079
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 990 gctccgctcttctccacgtactcgcgtacgaagaacgtcacgcatcttcgcagagat 1049
OY 1080 gcaagtcg-----tttcaagctgcggagatgatccgcggtgtggtctccggtgtgt 1130
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1050 tcaagcttgcatgagatgaacatgttgaattgttctcgaaggtgtgtctccggttgg 1109
OY 1131 ttaacgaagaagaacggtcgcgaagattcgggatttaagaacaatacgcgaagaattcacc 1190
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1110 tactgtataacggaataaacccaagcttggaagtattcctaagtaacgcggagatgagtc 1169
OY 1191 gaacattaaagtgtatgtatgtctcagattgtcttgccttccatctctggaacggttggga 1250
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1170 ggaagtgaattgttccgaagttcccttgattcaatatcatcaccgatttaabgttggga 1229
OY 1251 agagccagaagaacagatgaagtcgtgtgtataggtctctgtatgactccaccgaacctaat 1310
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 tgaagatgatagaagaacggtcgttttgaattgcaaccgaatatataltgcatgtaacatc 1289
OY 1311 ttccaacggtctgacgagaatctcaagagtgtccgtctgtaaatccgctgaattcaca 1370
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 tttaagaagaatgagatcgtgttc---atgcttgtgtggaagaagtgaagaatcgatctcgt 1346
OY 1371 aacgggtgaatcaactcgcgtccgtacatctccaacgaagaataccaacgaactcga 1430
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1347 caacgggattgtcgagacgtcatccgactcgcagag-----aatctcga 1391
OY 1431 agcaggaatggtcacaagaacaatcgtcgcgttaaaccaattcgtctacttgcttt 1490
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 1392 ttctcgttgataaaccggtttctccggagatgtgaacggtacggttaccggtcgcat 1451
OY 1491 agccgagccgttgccctaaaggtctcaagattcgtctaagttgacttcaactcaggaagaat 1550
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1452 tggagatcccgatgcccgaagaatctccggtgtgttgaaagcttgatgtgtctaaaggagatcg 1511
OY 1551 taagaacat-----cttaccggtataacggttaccgaaggagagacctct 1595
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1512 ggaatgttgacggttgcccgtaagaatgtacggttcaagttgtttaocggggaggaacctgt 1571
OY 1596 gtttctcccggaaga-----aggagagaaggaagacgaagaataatcctctg 1643
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1572 ttctgtaagtaaggatccctgttaalcctggaaggcggagagatgtatgtatgtgtgac 1631
OY 1644 ttctcgttcacgacgaagaacatactggaatcggagattacagatagtttaacgcgttgaact 1703
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1632 gtatgttcacgatgaaggtacgtcggaatccgaagttccggtgtgatggaacgccaatcgcc 1691
OY 1704 agaggttga-----gcaacggttaaaccttccgtcaaaaggttccgttacgaattcaacg 1757
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1692 ggaagcttgaatcgttcgcgcgctgaggtgtgcgcggaaggttccgttacgattccatgtg 1751
OY 1758 taatcattcggagccgcatgatttggcgaagc 1789
    || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1752 gtatttgcagaagaaagtgaccttaataagc 1783

```

RESULT 13

AAE77206 standard; CDNA; 1950 BP.

AAE77206;

16-MAY-2001 (first entry)

CDNA encoding sunflower neoxanthin cleavage enzyme (NCE).

Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;

NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;

pathogen resistance; abscisic acid metabolism; ss.

Hellianthus annuus.

Key Location/Qualifiers

CDS 1..1632 /tag= a

FT /product= "NCE"

FT /note= "Neoxanthin cleavage enzyme"

FT /partial

WO200112801-A2.

22-FEB-2001.

17-AUG-2000; 2000MO-US22961.

18-AUG-1999; 99US-0149656.

23-MAY-2000; 2000US-0206405.

(PION-) PIONEER HI-BRED INT INC.

(CDNA-) CDRAGEN CORP.

Bidney DL, Crasta OR, Hu X, Lu G;

WPI: 2001-211215/21.

P-PSDB: AAB72303.

Novel isolated defence-related signalling gene isolated from sunflower

encoding neoxanthin cleavage enzyme, amino acid permease or glutamic

acid-rich protein useful for increasing resistance of plant to a

pathogen

Claim 1; Page 94-97; 135pp; English.

XX This invention relates to defence-related signalling genes isolated from
 CC the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
 CC enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
 CC protein (GR). The signalling gene is useful for increasing the
 CC resistance of a plant to a pathogen such as fungus, virus, bacterium,
 CC nematode or insect (e.g. European corn borer), preferably
 CC *Sclerotinia* spp., *Phoma* spp., or *Phomopsis* spp., by stably incorporating a
 CC construct containing the gene into the genome of the plant. The gene is
 CC useful for regulating gene expression in a plant, in response to a
 CC stimulus such as infection with a pathogen, damage from a pathogen,
 CC hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
 CC oxalic acid or expression of a gene encoding oxalic acid oxidase. The
 CC genes are also useful for stem-preferred regulation of gene expression in
 CC a plant. The genes are useful in agriculture, particularly in the
 CC breeding of crop plants with improved agronomic traits, for modifying
 CC abscisic acid (ABA) metabolism and for modifying amino acid transport and
 CC content in plants. The present sequence represents cDNA encoding the
 CC sunflower neoxanthin cleavage enzyme (NCE).

SO Sequence 1950 BP; 535 A; 456 C; 453 G; 506 T; 0 other;

Query Match 9.4%; Score 169.2; DB 22; Length 1950;
 Best Local Similarity 51.0%; Pred. NO. 9.2e-44;
 Matches 480; Conservative 0; Mismatches 453; Indels 9; Gaps 3;

QY 347 cggctgattcctagtgatcagatcgccggaatttgcctccggtgaatgaagaacccgctcc 406
 DB 308 cagcttgatcccaaaacagcttctgtatgaacttccaacggtgagacactccctccga 367
 QY 407 ggcgtatcttcggtggtcggaactcccgatccatcaagaagatgtagtcgca 466
 DB 368 ctgactgtgaagatcatcgagcgacactgcccgaagtgccttcgaacggtgcttaacttcogta 427
 QY 467 acggaactaacccaactccaagcggtagacaggtcaccactcttcgaacggaacggtta 526
 DB 428 atggtccgaacccgaattctcttcgcgagagaccatccactctgtagtgagtagca 487
 QY 527 tgggtcagccggtcaaatctgaaacggttcagtagtagtagcttgcgggtttactcga 586
 DB 488 tgcctcatgcatctcgatcccaatggaagactcgtatgtagccgaatacacaataa 547
 QY 587 ctacacgggttgttcaggaactgaatgtgtgcacgggtttccccaagaacatcggtg 646
 DB 548 catacaaatatcaatagagaagaagcggtatcccatattccacaacgtgttttag 607
 QY 647 agcttcacggcacacacgggtatgcccgaatcgtatctc--taagccagaagctgacg 703
 DB 608 gtttaatgtgtgactgctctgcagctgcagatgcagtcacgtgcgcccattttg 667
 QY 704 ccggtatagtcgacccgacaggaacgggtgtagcttaacgcccgtttgttctatttca 763
 DB 668 ctgacaatttgaccaccaaaaggatgtgtcgaataccagatctggtcccttttg 727
 QY 764 atggccggttatgtgctatgcgagagatgattacttaaccaagtttagtcaactccca 823
 DB 728 gcaacgaacttttgcctcgtggaagtcggtatcccatatgcccgtcaactagcggccg 787
 QY 824 atggaagttaaaaaacggtgtgtcgatttgaattgatacaattagaatcacaatga 883
 DB 788 acggtgacatagtcacccgctgcagctgagacttcgacgcaaaactattcatgacatga 847
 QY 884 ttgcccacccgaagaagtgcacccgaatcgggtgaactcttgcctttaagctagacgtcgc 943
 DB 848 ccgctcaccaaaataatgatccatlaaagaagaagctttgtcctttgtagtgcacag 907
 QY 944 tttaaaagccttaactaaatcactccgattcctcaaccggaaggaactaatcacccagcg 1003
 DB 908 tcgcc---ccttcttaaccttttgcgtttcaacgaaaagggagaaaacaaacgcggtg 964
 QY 1004 tcgagatcagc---ttgatcagccaacgatgatgcagatttcgcgatttcagagaact 1060

DB 965 tcccgatcttccaatgacgaagcccgctgtcttccacgacttcgcacccaact 1024
 QY 1061 tcgtcgtgtacccctgaccagaagtcgtttccaagctgcggagatgtagtcgggtggt 1120
 DB 1025 acgcaatcttcccgagatcccaatcggagatgagccaaatggaatgctggtggtggtat 1084
 QY 1121 ctccggtgttcaagcacagaagcagagtcgaagatctggagatttgaacaatacggcg 1180
 DB 1085 ccccggttagcgcgagcgtggaagtgctcgtcgtcgggtgtagtccctcgtgaacgga 1144
 QY 1181 aagatcatcgacaactaagtatgtagtgcctccaagattgcttgccttcacatctcga 1240
 DB 1145 aagacgagtcagagatgagatggtttgaggttcgggttttaatgtagatcatgcatca 1204
 QY 1241 acggttggaagagccagaacaagatggaatgcgtcgtatag 1282
 DB 1205 atgcattggaagagatggtgcgagatcagtcgtgtagtgg 1246

RESULT 14

AAC57157
 ID AAC57157 standard; DNA; 491 BP.

AC AAC57157;
 DT 25-JAN-2001 (first entry)

DE Pinus radiata transcription factor DNA sequence #603.

KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB, ss.

OS Pinus radiata.

PN MO200053724-A2.

PD 14-SEP-2000.

PE 09-MAR-2000; 2000WO-US06112.

PR 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

PI Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI: 2000-579369/54.

DR New isolated polynucleotide encoding a plant transcription factor for

PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,

PT having modified gene expression or modified activity of a polypeptide

PT -

XX Claim 1; Page 598; 747pp; English.

FS The present invention relates to novel plant transcription factors from
 XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.

SQ Sequence 491 BP; 86 A; 143 C; 171 G; 91 T; 0 other;

Query Match 8.3%; Score 149.4; DB 21; Length 491;
Best Local Similarity 59.2%; Pred. No. 1.1e-37;
Matches 255; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```
QY 436 cccattccatcaaaagggtatgtgcaacagagtaaacccattcacgagccggtg 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 ccgagagtgcttagatgagtggtttacgtccgaatgscggaatccccgttcaaacccgc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 496 acaagaccacattcttcgaaggaacggtatgttcacgcccgtcaatttcgaacacggt 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 ggcggccacattatttcgacgscgagtgaatgatactgcgtgacgtcgaagacacggg 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 556 tcagcttaactcgcgttcggttcctacgaactaacgggttgcctcagaagaatcaattg 615
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 aagctagttacagttcgcgttcacggaagaccgaagagctcgttaagcgaagagcgggcg 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 616 ggtcgaaccggtttccccaagcactcgtgagcttcacagcgccacacggatltgccga 675
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 ggcggagagcttttaccgaagcccatcggaactccacggcgaagcggtggtgctgcgc 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 676 ctcaatgctattctacgcaagagctgcagccggtatagtcgaccggcacaacggaacgggt 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ctgctgcgtcagatggtgcggcggtctgcgtgcgttcacacacgggaagggcactgggc 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 726 gtagctaacgacgggttgcctatttcaattcagcggttatgttcggaagatgat 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 gtggctaatgcccggcggtgcctctttaaagcgctgcgtcgtcgtatgttcgaagaagcat 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 ttaccttaacgaatcagatcactcccaatgtagatlttaaaaaacggctggtcgttcgat 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 ctcccgatagccgtcaggtgtagcggtgtagcgcgatctgtgtagcaagcggtcggttcgat 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 ttgatggaca 866
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ttcgacggggca 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

AAC57162 ID AAC57162 standard; DNA; 386 BP.

XX AAC57162;

XX AC 25-JAN-2001 (first entry)

XX DT Pinus radiata transcription factor DNA sequence #608.

XX DE Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

XX KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;

XX KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;

XX KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;

XX KW type 2 Cys2His2; CCAAT box element; MYB; ss.

XX OS Pinus radiata.

XX OS WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Wood M, McGrath A, Shenk MA, Glenn M;

XX DR WPI; 2000-579369/54.

```
XX New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PS -
PS Claim 1; Page 599; 747pp; English.
```

```
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper, LIM domain, AP2
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
```

SQ Sequence 386 BP; 89 A; 105 C; 106 G; 86 T; 0 other;

Query Match 8.0%; Score 143.4; DB 21; Length 386;
Best Local Similarity 70.0%; Pred. No. 8.4e-36;
Matches 208; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

```
QY 1103 agatgacgcgggtggtcctcgggtgtttagcagaagaacaggttcgaagattcggga 1162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 aaatgatacagaaggcgtctcctccagttattcaacaagaagaaggctccgcgttcgggc 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1163 tttagacaatacgcgcgaagatltcatcgaacattgaatgattgattcctcgaattgct 1222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 ttctgcccaataatgctcttcgacgagtgatgagctgaatgatacgaggtcccgattgct 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1223 tctgttccatctctcgaaacgcttgggaagagcagaacagatgaagtctgctgataag 1282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 tctgtttcatctctcgaaacgcttgggaagagcagaacagatgaagtctgctgataag 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1283 gttcctgtatgactcaccagagctcaatttccaagagcttgacgaagatctccaagagtg 1342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 178 gttcctgtatgactcaccagagctcaatttccaagagcttgacgaagatctccaagagtg 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1343 tctgttctgaataatccgcttgaatctcaaaacgggtgaatcgaactcgcgtccgatca 1399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 238 ttctgtcggaattcgtgctcaatctcaaaacgggttctccaccagacgagatca 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 24, 2002, 06:25:34
Job time: 6890 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 05:14:15 ; Search time 1694.89 Seconds
(Without alignments)
14333.980 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atgcgtcttcacgcgaac.....tggcgaagcagctcgtgtga 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hlc: *
9: gb_estl: *
10: gb_estl2: *
11: gb_hlc: *
12: gb_gss: *
13: em_gss_hum: *
14: em_gss_inv: *
15: em_gss_pln: *
16: em_gss_vtc: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	425.6	23.6	720	10	BM412731 EST587047
2	417	23.2	805	10	BM408615 EST582942
3	416.2	23.1	781	12	BH549344 BGR053TF
4	382.8	21.3	643	9	AW933524 EST359283
5	375.6	20.9	787	10	BM408565 EST582892
6	344	19.1	592	10	BF113346 EST440936
7	332.4	18.5	566	10	BE461924 EST413439
8	323.2	18.0	553	10	BM085672 sa128a02.
9	318	17.7	559	10	BM085005 sa131a08.
10	316.6	17.6	520	10	BE432853 EST399478
11	313.8	17.4	742	12	BH559796 BOHR047TF
12	312.8	17.4	594	12	BH458011 BOHR02TF
13	309.4	17.2	547	10	BI974879 sa174b11.
14	308.6	17.1	547	10	BE437072 EST408190
15	303.4	16.9	617	10	BE458861 EST414153
16	300.4	16.7	546	10	BM084948 sa130b08.
17	295.8	16.4	618	9	AW443298 EST308228

18	294	16.3	564	10	BM085488 sa137d09.
19	292.8	16.3	627	10	BF051297
20	289.2	16.1	509	10	BE451573
21	288.6	16.0	495	10	BE434930
22	276.2	15.3	490	10	BF050563
23	275.4	15.3	657	9	AA556214
24	267.4	14.9	503	12	B27476
25	254.8	14.2	566	10	BE459895
26	249.2	13.8	450	10	AW930245 EST340702
27	248	13.8	552	10	BG039692 NXSL102.
28	245	13.6	778	12	BH579654
29	236.8	13.2	600	12	BH458018 BOHR02TR
30	233.2	13.0	424	12	BH496327 BOH012TF
31	224.4	12.5	337	12	BH579641 BGR073TF
32	220.4	12.2	410	10	BE525530 M26D7STM
33	217	12.1	446	10	BE471089
34	216.6	12.0	713	10	BI928174
35	209.6	11.6	495	9	AU084510
36	207	11.5	502	12	CNS000RV
37	205.6	11.4	590	10	BG592987
38	203.2	11.3	447	9	AW933245
39	202.8	11.3	492	9	AU084760
40	201.4	11.2	412	9	AV419581
41	195.2	10.8	680	12	AQ969739
42	185.8	10.3	549	10	BE425834
43	173.6	9.6	500	9	AW289745
44	164.6	9.1	678	12	BH593312
45	152.6	8.3	267	10	BE529289

ALIGNMENTS

RESULT 1
BM412731 720 bp mRNA linear EST 22-JAN-2002
LOCUS
DEFINITION EST587047 tomato breaker fruit Lycopersicon esculentum cDNA clone
CLEG60N24 5' end, mRNA sequence.

ACCESSION
BM412731
VERSION
BM412731.1 GI:18264350
KEYWORDS
EST.

ORGANISM

Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE
1 (bases 1 to 720)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai

AUTHORS
'J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CWGI

JOURNAL
Clemson University Genomics Institute

COMMENT
100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics

Institute
Seq primer: T3.

FEATURES

source
Location/Qualifiers
1..720
/organism="Lycopersicon esculentum"
/cultivar="T496"
/db_xref="taxon:4081"
/clone="CLEG60N24"
/clone_lib="tomato breaker fruit"
/issue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSMCadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;

[illegible]

OY		334	ccgccttcctaataagcgcgtgatccctagtgttcaagatcgccggaattttgctccggtgaat	393
Db		61	cctctttccccaaaacggctgatctcacgttatTTCAAATCGCCGGGAATTATTTCCCGGTCGC	120
OY		394	gaacagcccgcctccggcgtaattcttcgggtgycgcggaaaacttcaccgattccaacaaga	453
Db		121	GAAATTTCCCCTCCGAGAGGTGTCTCGACGCTTGGAAGGACATCATCCTTACGTCAATTAA	180
OY		454	gttgatgtgpcgaacgagagctaacaccttccacgagccggtgacagagtlaccacctcttc	513
Db		181	GTTTATGTGCCGTAACGGTGTGCAAAATCCGATGTTTCGAGCCAAATCGTCGGGCATCATTTGTTTC	240
OY		514	gaacgagagcgcgtatagttttagcgcgcgttcaaatctcgacaacgcgttcagctagctcgttc	573
Db		241	GACGAGAGAGGGAATGGTTTCACGCACGTTTAAAATAAACCAAGGTTTCAGCAAGCTACGCATATC	300
OY		574	cgatttacccaactcaacccggttgttccacgagaagctcaatttggctgcagccggttttcccc	633
Db		301	CGGTTTACTAAAAACCGAGAGATTATTTCAAGAAAAAGATTGGGTTCGACCGGTTTTTTCG	360
OY		634	aaagcatcgcgtgtgagcttcaacgacacacacgcggtlatgtccgcacatgctatttcaacgcc	693
Db		361	AAAGCAATCGCGGACCTTCACGGCCACTGGGAATCGCACGTCGTATGCTTTTAAACGA	420
OY		694	agagctgcagccgcggtatagtcgaccgcgcacacgcggaaccgcggtgtgactgaacgcggttg	753
Db		421	CCTGGGCTTTGGGGCTCTCTCAACAATCAAAAGGGCGCTGGGGGTATGACCAACCCCGTTTGG	480
OY		754	gctatttcaatcgcgcgcgttatgtgctatgtcgcgagatgattcaaccttaaccaagtctcag	813
Db		481	GTTTACTTCAATAAACCGGCTTTTAGCAATGTCAAGAAGACGATTTCACGTACCAATTAAAA	540
OY		814	atccactcccaatlgsgagatttaaaccgcgttggctggcttgcatttgcattgagacaattaga	873
Db		541	ATAACTCAAAACCGGACGACCTCCAAACCGCTGGGCGCTTACGATTTTCGACGCTCAATTAAA	600
OY		874	tccacaatgatgtgcccaaccgcgaagaatcgacccgcggaatccgcgtggaactctgtccttaagc	933
Db		601	TGATGATGATGTGCCACCACCGAAGCTCGACCCGGTTACAAAGGACCTCCACCGCTTAACG	660
OY		934	tacgacgcctgcattcaaaagccttcaataatacttcgcatttccacgcgagcgaactaa	993
Db		661	TACGACGTCGTAAAAAAGCTTACTGAAATTTTTCAGATTCTCCGCGGACGCGCTTAAA	720
OY		994	tcaacgcgcctgcagatctgaattcagcttgatcaacgacacagatgacgagatttcgcgattaca	1053
Db		721	TCACCGGAGACTGGAATCCCGCTCGAGACTCCGAGATGATGTTACAGATTTCCTATTAACG	780
OY		1054	g 1054	
Db		781	G 781	

RESULT 4
AM933524 LOCUS AM933524 643 bp mRNA linear EST 18-MAY-2001
DEFINITION CDNA clone cIEF54E2 5', mRNA sequence.
ACCESSION AM933524
VERSION AM933524.2 GI:11388127
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE 1 (bases 1 to 643)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Updon,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Rongning,C.M.,
Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.,
Generation of ESTs from tomato fruit tissue

JOURNAL Unpublished (1999)
On May 30, 2000 this sequence version replaced gi:8108841.

COMMENT
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1..643
Location/Qualifiers

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF54E2"
/clone.lib="tomato fruit mature green, TAMU"
/tissue.type="fruit pericarp"
/dev_stage="mature green (3-5 days pre-ripening)"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; CLEF - Fruit were tagged at the 1cm stage and harvested 3-5 days prior to ripening. Fruit were cut in half to verify the seeds were indeed 'immature' and the seeds and locules were discarded prior to freezing the pericarp."
130 c 142 g 186 t 1 others

BASE COUNT 184 a 130 c 142 g 186 t 1 others
ORIGIN

Query Match 21.3%; Score 382.8; DB 9; Length 643;
Best Local Similarity 74.7%; Pred. No. 3.8e-103;
Matches 480; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 685 ttctacccagagctgcagcggatagctgcagccgacagcaacccggtgtagctaac 744
DB 1 TTATTAGCGTCGTGGCTCTTCGACCTGTGATCACAGTAAAGAACTGGTTGCAAC 60
QY 745 gccggttctgctattcaatgagcgggtatggtatgctgcgaggaatgattaccctac 804
DB 61 GCCGGTTAGTGTTCATTAACCCGATCTGCTATGCTGAAGATGATTGCCCTTAC 120
QY 805 caagttcagatcctcccaatgagatttaaaacgggtgctggttcgatttctgtaga 864
DB 121 CATGTAAAGGTAACCCACCGCGAGCTTAAACAGAGGCTGATTCGATTTCGAGGC 180
QY 865 caatagaatcccaatgattgcccacccggaagtcgacccggaatccggtgtaacttc 924
DB 181 CACTTAAATCCACATGATGATGACCAACCAAGCTGACCCAGTTCCGGTGAGCTATT 240
QY 925 gctttaagctagcagctgcttcaaaagccttaacattcgcgattcctcaccgag 984
DB 241 GCTCTTAGCTACGATGTGATTCAGAAAGCCATACCTCAAGTACTTCAGATTTCAAAAT 300
QY 985 ggaactaaatcacccagcgtcgagattcagctgattcagcagcaagatgagcagattc 1044
DB 301 GGGGAAAATCAATAATGTTGAAATTCAGTTGAAGAACCAACAAATGATGATTC 360
QY 1045 ggcattacagagaaactcgctgcgtacgtacacgaagtcgcttcttcaagcgcggag 1104
DB 361 GCATTAAGTACGAACTTCGTCGATTCCTCGATCAACAAATCTCTTTCAAGATGTCGAA 420
QY 1105 atgataccggtggtgtcctcggtgtttagcagaagaacaagtcgcaagatcggatt 1164
DB 421 ATGATCCGTGAGAGTTACCGGCTGTTACGACAAGAAAGTTCCCGATTTGGATTC 480
QY 1165 ttgagacaataagccgaagattcatcgaaatgaagtgtattgattcgaagttcttc 1224
DB 481 CTGGATTAAGTACCGGAAAGTGGCTGTGATTAAATGGGTTGAAGTACCTGATTTGTTTC 540
QY 1225 tgcctccatctctggaagcttgggaagagccggaacaagaatgaagtcgctcgatagg 1284
DB 541 TGTTCACACTCTTGGAATGCTTGGGAAGAGCAAGAAACAGATCAATCTGTTGTAATTGTT 600
QY 1285 tccgtatgactcaccagactcaatttcaacgagtcgtacg 1327

DB 601 TCATGTATGACACCACGACTCCATTTTCATGATGATGATG 643

RESULT 5
BM408565 787 bp mRNA linear EST 22-JAN-2002
LOCUS EST582892 tomato breaker fruit Lycopersicon esculentum cDNA clone
DEFINITION CLEF45E23 5' end, mRNA sequence.
ACCESSION BM408565
VERSION BM408565.1 GI:18260195
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE
1 (bases 1 to 787)
Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
J., Bougri,O., Kirkness,E., Uterback,T., Van Aken,S., Renning
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato fruit tissue, breaker stage (2002)
Unpublished (2002)
Contact: CUGI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.

FEATURES
source
1..787
Location/Qualifiers

/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEF45E23"
/clone.lib="tomato breaker fruit"
/tissue.type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"

/note="Vector: pBluescriptSKmCunadapt; Site_1: EcoRI;
Site_2: XhoI; supplier: Boyce Thompson Institute;
sequencing: The Institute for Genomic Research. Fruit
were harvested at the breaker stage (first sign of
lycopene accumulation on the blossom end of fruit). Fruit
were cut in half and the seeds and locules were discarded
prior to freezing the pericarp."

BASE COUNT 216 a 194 c 165 g 212 t
ORIGIN

Query Match 20.9%; Score 375.6; DB 10; Length 787;
Best Local Similarity 67.8%; Pred. No. 6.2e-101;
Matches 525; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

QY 205 tctcccgcaattgttcttaagcccaagccaagaatccaacctaagaatgattg 264
DB 5 TCACACCCAAACAGAAACAACACTCTCTTCTTCAACTTCACAGTGGAAATTGA 64
QY 265 ttccagagacgagcgcgcgcgcgttgcagcgcgcgcgcgttctcttgcacacagag 324
DB 65 GTGCAGAAAGCAGACAGCAATAGGCTTTAGATGCTGTAGAAAGTCTTTAACTAAACATGAA 124
QY 325 aagctaacccgcctctcctaagcgtgacactagtggttcagatcgccggaatttgct 384
DB 125 CTGGAACACCTTTGCCGAAACACGCCGAGATCCAGATTTCTGGGAATTTTGGT 184
QY 385 ccggtgaatgaagaccgcgcgcgcgttaattcctcggttggtcggaanaacttcccgcttc 444
DB 185 CCGGTACCGGAAATCCAGTCTGTCAATCTTCCGCTCACCGGAAATAATCCCAATATGT 244

OY	445	atcaaaagagtgatgctgcgaacggaaactaacccacttcaaggacgggtgacagttaac	504
Db	245	gTTTCAAGGCGTTTTACGTTTCGAAGCCGAGACTTAACCTCTCTTTTGGAAACCAACCCGCCGACAC	304
OY	505	cactccttcgcagcgagacgcatgattgtttacagccgtccaatttcgaacacaggttcagctaac	564
Db	305	CATTCTTCGCACGGCGACGGATATGGTTTACGCCGTTCATTTCAAAAATATGGTGCGCTACT	364
OY	565	tacgcttcgcggtttactcacgactaacaccggttgttcagaagcacaattgtgttcgacgc	624
Db	365	TACGCTTCGCCGTTTCACATGAAACAAGAGGCGCTGTGTTCAAAGAAAAAGCTTTGGGTGGCCT	424
OY	625	gtttcccccaaaagcatalcgtgtgtagcttcaagggcacaacggatattgcccgactatgcta	684
Db	425	GTTTTCCTCTAAAGCCATTTGTGTGAATTAATATGGTCACTCTGTGAATTGCAAGCGTTATGCTG	484
OY	685	ttctacgcgcagagcgtcgacgcgggtatagtcgacccggcacaacggnaacccggtgtagctaac	744
Db	485	TTTTACGCTCGGGGGCTCTTGCGACTTGTGTATCACAGTAAAGGAACGTGGTTTCCAAAC	544
OY	745	gccggtttgcgtatattcaatlbgccgggttatgtgatgctatgctcgagaatattacctaac	804
Db	545	GCCGGTTTAGTCCTATTTCATATACCAGATTACTGTGTATGTCTGAGAATCATTTGGCTTAC	604
OY	805	caagtcagatcacctcccgaatygagattaaaaaccggttgcgtgcgatttgatygta	864
Db	605	CATGTAAGAGTAACACCCACCGCGCATCTTTAAACAGAGGGTCGATTCGATTTCAGCGGC	664
OY	865	caattagaatccacacatgatgtgcccacccggaaagtcgcagccggaaatcgggtgaactcttc	924
Db	665	CAGCTAATAATCCACCATGATATAGCTTACCACCAAGCTCGACCCGATTTCCGGTGAAGTATTT	724
OY	925	gcttaaagcttcagcagctgcgtttcagaagccttcataaatattcttcagattcca	978
Db	725	GCTCTACTTACGATGTGATTCAGAAAGCCATPACCTCCAGTACTTACGATTTCATTTCA	778
RESULT	6		
Bf113346			
LOCUS	Bf113346	592 bp	mRNA linear EST 18-MAY-2001
DEFINITION	EST440936 tomato breaker fruit Lycopersicon esculentum cDNA clone		
ACCESSION	CLEG44B9	5' sequence,	mRNA sequence.
VERSION	Bf113346		
KEYWORDS	Bf113346.1 GI:10943036		
SOURCE	EST.		
ORGANISM	tomato.		
	Lycopersicon esculentum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;		
	Lycopersicon.		
REFERENCE	1 (bases 1 to 592)		
AUTHORS	Alcala,J., Vredelov,J., White,R., van der Hoeven,R.S., Holt,I.E., Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronging,C.M., Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley ,S.D. Generation of ESTs from tomato fruit tissue, breaker stage unpublished (2000) Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html. Location/Qualifiers 1..592 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="CLEG44B9" /clone_lib="tomato breaker fruit" /tissue_type="Pericarp" /dev_stage="breaker" /lab_host="SOLR"		
TITLE	JOURNL		
COMMENT			
FEATURES	Source		

/note="vector: phluescriptsmcTadapT, Site1: EORI; Site2: XhoI, supplier: Boyce Thompson Institute; sequencing: The institute for Genomic Research. Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of fruit). Fruits were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT	194 a	106 c	131 g	161 t
ORIGIN				
Query Match	19.1%	Score 344;	DB 10;	Length 592;
Best Local Similarity	73.8%;	Pred. No. 1.5e-91;		
Matches 437; Conservative	0;	Mismatches 155;	Indels 0;	Gaps 0;
OY	893	cgaagtcgaccgcgaatccggtgacatcttgctttaagcagcagcgtcgcttcaagc	952	
DB	1	CAAACTCGACCAATGTTCCGGTGACTATTGGCTTTAGCTACGATGTGATTCCAGAAC	60	
OY	953	cttaccctaaatctctcgcattctcaccgcgagcaactaatcacgcgacgctcgatc	1012	
DB	61	CATTACTCAAGTACTTCAGATTTCAAAAAATGGGAAAAAATCAATGATGTTGAATTC	120	
OY	1013	agcttgatcagcgaacgatgatgcagatctcgcgattacagaacacttgctcgtaac	1072	
DB	121	CAGTTGAAGACCCCAATGATGATGATTTGCAATTTCTGAGAACTTCGTCGTCATTC	180	
OY	1073	ctgaccagcgaatcgcttttcaagctcgccggaatgatccgcggtggtctccggtgtt	1132	
DB	181	CTGATCAACAACTGCTTTTCAGATGTCGAAATGATCCGTGAGAGTTCCACCGTGCTTT	240	
OY	1133	acgacaagacaagctcgcaagatccggattctagaacaatacgcgcgaagatcatcga	1192	
DB	241	ACGACAAGACAAAGTTTCCGATTGGTATCTCGATATGATACCGGAAGATGGGTCG	300	
OY	1193	acattaatgatgatgatgctccagatltgctctgcttccatctctggaagcttgggaag	1252	
DB	301	ATTGAAATGGGTTGAAGTACTGATTTGTTCTGTTTCCACCTCGAATGCTTGGGAAG	360	
OY	1253	agccgaagaacagatgaatgctgcgcatgatagggttcctgtatgatctcacagatcaatt	1312	
DB	361	AAGCAAGAAACAATGTAATCGTTGAATGGTTCATGATGACACCACCAAGCTCCATT	420	
OY	1313	tcaacgagctctacagaaatctcaagagtgctcgtctcgaatccgcctgaatccaaga	1372	
DB	421	TCAATGAATGTTATGAAGGCTTAAGAGTGTTTTATCCAAATTCGCTCAATTGGAAAA	480	
OY	1373	ccggtgaatcaactcgccgctcgcatcatctccaacggaatcaacgaagtcaactcoga	1432	
DB	481	CAGGGAATCAACAAGAAATCCATATGCAAAACCCGGATGAACAGATTTAAGAAG	540	
OY	1433	caggagatggtcaacagaacatgcttcgcccgcgctaaacccaattgcttact	1484	
DB	541	CTGGAATGGTGAACCGAAACAATCTCGAAGAGAAAAACAGATGATGCTTATTT	592	
RESULT	7			
LOCUS	BE461924	566 bp	mRNA	linear
DEFINITION	EST1413439 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA			
ACCESSION	BE461924			
VERSION	BE461924.1	GI:9506322		
KEYWORDS	EST.			
SOURCE	tomato.			
ORGANISM	Lycopersicon esculentum			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
AUTHORS	Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,			
	Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,			
	1 (bases 1 to 566)			

Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
 TITLE Generation of ESTs from tomato fruit tissue, breaker stage
 JOURNAL Unpublished (2000)
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.

FEATURES

SOURCE

Location/Qualifiers

1..566

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db.xref="taxon:4081"

/clone="cLEG40L2"

/clone_lib="tomato breaker fruit, TIGR"

/tissue_type="pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/note="Vector: pBluescriptSKmCudaPt; Site1: EcoRI; Site2: XhoI; Fruit were harvested at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit). Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 184 a 104 c 124 g 154 t

ORIGIN

Query Match 18.5%; Score 332.4; DB 10; Length 566;
 Best Local Similarity 74.2%; Pred. No. 4.3e-85;
 Matches 420; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

893 cgaagctgacccggaatccggtgacatcttccttaagcagcagcgtcttcaaacg 952
 1 CAAGCTGACCCAGTTTCGGTGAAGTATTTGCTTTAGCTACGATGATTCGAAAC 60
 953 cttaacttaaatctccgattctcaccgagcgaactaacacgagcgtcgatc 1012
 61 CATACCTCAAGTACTTCAGATTTTCAAAAATGGGAAATCAATGATGTGAATTTC 120
 1013 agcttgatcagcgaacgatgatgcagatctcgagattacagaactcgctcgatc 1072
 121 CAGTTGAAGACCAACAATGATGATGATTCGCAATTACTGAGAACTCGTCATTC 180
 1073 ctgaccaggaatcgcttcttaagctcgccgagatgatccggtggtctcgatgatt 1132
 181 CTGATCAACAGTCGTTTCAAGATGTCGAAATGATCCGTGAGGTTACCGGTGTTT 240
 1133 acgacaagaagaagtcgcaagatcgagatttagacaatacgcggaagattcatcga 1192
 241 ACGACAAGAACAAAGTTTCCGATTTGGTATCTGATTAAGTACCGGAAGATGGGTGC 300
 1193 acattaatgatgatgatccagattgcttctgcttccattcttggaagcgttggaag 1252
 301 AATTAAATGGGTTGAAGTACCTGATTTGTTCTGTTCCACCTCGAATCGTTGGGAAG 360
 1253 acccgaaacagatgaagctcgtcgatgaggtcctgtatgactccaccagctcaatt 1312
 361 AAGCGAACAACATGAATCGTTGTAATGGTTCAATGATGACACACGACTCCATT 420
 1313 tcaagagctcgaagaaatcgaagatgtcgtctgtaaatccgcgtgaatcaca 1372
 421 TCAATGATGTGATGAAGGCTTAAGAGTGTATTATCCGAATCCGTCTCAATTGAAAA 480
 1373 ccggtgaatcaactcgccgtcgatcatctccaacgaagatcaacgaatcgaag 1432
 481 CAGGGAATCAACAAGAAATCCATATCGAAGAACCCGGATGAACAAGTATTAAGAG 540
 1433 caggatggtcaacagaacatgctc 1458
 541 CTGGAATGTTGAACCGAAACAAATC 566

RESULT 8
 BM085672 553 bp mRNA linear EST 19-NOV-2001
 LOCUS 3aj28a02.y1 Gm-cl066 glycine max cDNA clone SOYBEAN CLONE ID:
 DEFINITION Gm-cl066-3867 5' similar to YR:024023 024023 NEOXANTHIN CLEAVAGE

ENZYME: mRNA sequence.

ACCESSION BM085672

VERSION BM085672.1 GI:16996300

KEYWORDS

SOURCE

ORGANISM

soybean.
 Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.

1 (bases 1 to 553)

REFERENCE

AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Expelding, J., Corvelli, V., Khanna
 A., Bolla, B., Maria, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
 Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
 Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
 R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
 R., Waterston, R. and Wilson, R.

Public Soybean EST Project
 Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project

TITLE

JOURNAL

COMMENT

Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact: cd@resgen.com web site:
www.resgen.com

High quality sequence stop: 426.

Location/Qualifiers
 1..553
 /organism="Glycine max"
 /db.xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl066-3867"
 /clone_lib="Gm-cl066"
 /tissue_type="leaf and shoot tip, salt stressed, 2 week
 old seedling"
 /lab_host="DH10B"

FEATURES

SOURCE

ORIGIN

Query Match 18.0%; Score 323.2; DB 10; Length 553;
 Best Local Similarity 74.1%; Pred. No. 2.4e-85;
 Matches 409; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

822 caatgagatttaaaacccgttgctcgatttgatgacatagatccacat 881
 1 CAACGGCAGCTTAACCGCTCGGCGCTTACACATTTAACGCGCAGTTAAATCCACAT 60

BASE COUNT 148 a 165 c 120 g 120 t

ORIGIN

1..553
 /organism="Glycine max"
 /db.xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl066-3867"
 /clone_lib="Gm-cl066"
 /tissue_type="leaf and shoot tip, salt stressed, 2 week
 old seedling"
 /lab_host="DH10B"

/note="Vector: pBluescript II SK+, Site1: EcoRI, Site2:
 XhoI; The cDNA library was constructed from mRNA isolated
 from unexpanded leaves and the shoot tips of 2 week old
 seedlings from the cultivar Williams. The 2 week old
 seedlings were salt stressed in a solution of 500mM NaCl
 for 3 days prior to harvesting. Complementary DNA was
 synthesized from mRNA using a primer consisting of a
 poly(dT) sequence with a XhoI restriction site. EcoRI
 adapters were ligated to the blunt-ended cDNA fragments
 followed by XhoI digestion. The cDNA fragments were
 directionally cloned into the EcoRI-XhoI restriction site
 of the pBluescript vector. The ligated cDNA fragments were
 transformed into DH10B host cells (GibcoBRL). This library
 was constructed in the laboratory of Dr. Randy
 Shoemaker."


```

RESULT 10
BE432853
LOCUS BE432853 520 bp mRNA linear EST 18-MAY-2001
DEFINITION Esn399478 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone CLEB1167, mRNA sequence.
ACCESSION BE432853
VERSION BE432853.1 GI:9430792
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 520)
Alcala,J., Vredalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES
Source
1..520
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEB1167"
/clone_1lb="tomato breaker fruit, TIGR"
/tissue_type="Pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/Note="Vector: pBluescriptSKmCudaapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."
BASE COUNT 156 a 100 c 116 g 148 t
ORIGIN
Query Match 17.6%; Score 316.6; DB 10; Length 520;
Best Local Similarity 75.9%; Pred. No. 2,2e+83;
Matches 391; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 829 gatttaaaacggttgctggttcgatttgatgacaaatagaaatcacaaatgatggc 888
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2 GATCTTAAACAGAGGCGATTCGATTCGACGGCCAGCTAAATCCACATGATAGCT 61
QY 889 caccggaagtcgaccgcggaatcgcgtgaactcttcgcttlaagctagcagctgttca 948
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 CACCCAAACGCTGACCCAGATTCGCGTAGCTATTGCTCTAGCTAGCATGATTCAG 121
QY 949 aagccttactaaataactcgcattctacccggaagcgaactcaataccagcagctcgag 1008
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 AAGCCATACCTCAAGTACTTCAGATTTCATAAAATGGGAAAAATCAATATGATTGA 181
QY 1009 attcagctgatacgcacacgatacgaattcgcgatacagagaactcgtcgcc 1068
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 ATTCCAGTTGAAGACCAACATGATGATTTCCGCAATTACGAGAACTTCGCGTC 241
QY 1069 gtacctgacagcaagctgcttccaagctgcggaatgataccgcggtgtccggtg 1128
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 ATTCTGATCAACAAGTCGTTTCAAGATGCTGAATATGATCCGAGAGGTTACACG 301
QY 1129 gtttcgacgaagaagaagtcgcaagatctcgagatttagaacaataagccgaagatca 1188
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 GTTTACGACAAAGCAAAAGTTTCCGATTTCGATTTCGATTAAGTACGGAAGATGGG 361

```

```

QY 1189 tcgaacatgaatgcatgatccagatgtcttcgtccatctctggaagcttgg 1248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 TCTGATTTGAATGGGTTCAAGTACTGATTTCTGTTCTGTTCCACTCTGGAATGCTTGG 421
QY 1249 gaagagccagaacagatgaatgctgctgataaggtctctgatactaccacagactca 1308
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 GAAGAAGCAGAAACAGATCAATTCGTTGTAATGGTTCATGATGACACCAACAGACTCC 481
QY 1309 atttcaacgagctcgacgagaatctcaagagctg 1343
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 ATTTCAATGATGATGATCAAGGCTTAAGACTGT 516
RESULT 11
BH559796
LOCUS BH559796 742 bp DNA linear GSS 14-DEC-2001
DEFINITION BOHRR47TF BOHR Brassica oleracea genomic clone BOHRR47, DNA
sequence.
ACCESSION BH559796
VERSION BH559796.1 GI:17811576
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 742)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
JOURNAL Other GSSs: BOHRR47TR
COMMENT Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: Sheared ends.
FEATURES
Source
1..742
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOHRR47"
/clone_1lb="BOHR"
/Note="Vector: pHOS1, site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT 190 a 235 c 117 g 200 t
ORIGIN
Query Match 17.4%; Score 313.8; DB 12; Length 742;
Best Local Similarity 84.1%; Pred. No. 1.9e+82;
Matches 354; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 1 atggtcttcttaagggcaagcgtcgtgttctggtggaatggtcgtgtgcaatactact 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 321 AAGGCTTTCTTCACGGCAGCTACGGCGTTCGGAATGGGTGGTGGTAATATATACT 380
QY 61 cagcgccatatacgtctctctcaagcgcgcagctgagttatgtagctcttactctg 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 381 AAACCAACCTTATTCGTTCTTCTAAAGCTCCGCTTGCGTTATTCCTCCGTACCCATG 440
QY 121 gccagctgtgtacacgaatgaatgcttcaatgttcatctgcgcttcaacactccagctct 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 441 ACAAAATCGTCCCAAGAAAGCTCAATGTTCTCTGCGCTTCACACATCATCCGCTCTC 500
QY 181 catctccctaagcaatcatcaaacctcccgccatctgttgttaagcccaagaagaa 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 501 CATTTCCCAACCAATCTCCACCTCTCCGCAATTGTTGTAACCCAAACCAAGAA 560

```

QY	241	tcgaactaaacagatgaattgttccagagagagcgcgcgagcggttggacgcg	300
Db	561	ttccgacacccaacacatgaactgtttccaaaagcgcgcgccgcttgacgcg	620
QY	301	gaggtttccctgttcaagcagcagaagtaaccgccgtcttccaaacgctatcctagt	360
Db	621	gaaggtttcttgtatgacgcagcagacagatcccttcccaaaaacgcccagctcttgc	680
QY	361	gttcagatccgcggaatttgcctcgggttgaatgaacagccgcctcggcgtaattccg	420
Db	681	gttcaaatccgcgcggaactcttcgctccgggtgaacgaacagcctctcgttaacctccc	740
QY	421	g 421	
Db	741	g 741	
RESULT 12			
BH458011/c			
LOCUS			
DEFINITION BOHBT02TF BOHB Brassica oleracea genomic clone BOHBT02, DNA			
ACCESSION			
VERSION BH458011			
KEYWORDS BH458011.1 GI:17643722			
SOURCE			
ORGANISM			
Brassica oleracea.			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.			
REFERENCE			
AUTHORS Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.			
TITLE Whole genome shotgun sequencing of Brassica oleracea			
JOURNAL Unpublished (2001)			
COMMENT			
Other-GSSs: BOHBT02TR			
Contact: Chris Town			
TIGR			
9712 Medical Center Drive, Rockville, MD 20850, USA.			
Tel: 301-838-3523			
Fax: 301-838-0208			
Email: cdtown@tigr.org			
DNA is from a doubled haploid provided by Tom Osborn.			
Seq primer: TF			
Class: sheared ends.			
FEATURES			
SOURCE			
location/Qualifiers			
1..594			
/organism="Brassica oleracea"			
/strain="T01000DH3"			
/db_xref="taxon:3712"			
/clone="BOHBT02"			
/clone_1fb="BOHB"			
/note="Vector: pHD1. Site_1: BstXI; 2-3 kb sheared			
genomic DNA inserted into pHD1 using BstXI linkers"			
BASE COUNT			
138 a 170 c 131 g 155 t			
ORIGIN			
Query Match 17.4%; Score 312.8; DB 12; Length 594;			
Matches 414; Conservative 0; Mismatches 137; Indels 9; Gaps			
1,1,1			
QY	1047	gattcagacagactcgtctcgtctacccgacagcagaagtcgtttcaagctcgagagat	1106
Db	594	gattcagacagacactctcgtctcgtctccacccagcagaagtcgtttcagctcggagat	535
QY	1107	gattcgcggttggtcttcctcgggtgtttcaagacaagaagtcgaagattcgggattt	1166
Db	534	gattcgcggttggtcttcctcgggtgtttcagcacaacgaagtcgaagattcgggattt	475
QY	1167	agacacataacgcggaagattcatcgacaactaagtgatgtatgtctccagattgcttctg	1226
Db	474	gagcgaatracgttgaaagccttctcgttcgattcgatcggtgacgtgagagactgcttctg	415

OY	1227	cttcacatctcttgyaaacgctttgggaagagccgcgaacaaacagatgaagtcgttcgcatagtggtc	1286
Db	414	TTTTCCATCTCTTGGAACGCTTGGGAAGAGCGGAGAACAAGACGAGTCGTCTATCGGGGTC	355
OY	1287	cgtgatgactccaccagactcaatttcacaagagtcgtacaggaatctccaagaagtctct	1346
Db	354	ATGCATGACGCCGCCGCGACTGCATTTTCACGACAACACGAAACACTTTCAGAGTCTTT	295
OY	1347	gtctgaatatccgctcgtgaatctcaaaaaccggtgtaatcaactcgccgtccgatctccaa	1406
Db	294	GTCGGAGATTAACATAACCTGAACCAACAGGAGATCCACGCTCGCGCGATTATCTCCGA	235
OY	1407	cgaagatacaacaagtcacacctgaaagccggagtggttcaacagaaaaaatgctgcgcgtaa	1466
Db	234	GAAAT-----TAAATCTCGAAGCCGGATGTGTAACCGGAATCTTTTAGGTAGTAAGAA	184
OY	1467	aaccaaatcgcctaacttgcttagtcagccgagccgtgyccttaagctcacagatctgcctaa	1526
Db	183	AACGGCGTTTCGGTTACCTTGCTTTAACGSAACCGGTGGCCTAAAGTGTCCGGTTTGGCGAA	124
OY	1527	agttgatactcaactcgtggaagttaaagaacatcttacggcgataaacggttaccgagag	1586
Db	123	AGTGCACTTATCTACCGCGAAGAGTTCAAAATAATATATACGTGTGACGGTAACGTGGAGG	64
OY	1587	agaagcctctgtttctcccg 1606	
Db	63	AGAGCCTCTGTTTATGCGG 44	
RESULT 13			
B1974879/c			
LOCUS			
DEFINITION	B1974879	547 bp	mRNA linear EST 30-NOV-2001
ID:	sa174db1.v1 Gm-cl068 glycine max cDNA clone GENOME SYSTEMS CLONE		
CLEAVAGE ENZYME:	ID: Gm-cl068-4294 5' similar to TR:O24023 O24023 NEOXANTHIN		
VERSION	B1974879		
KEYWORDS	B1974879.1 GI:16349284		
SOURCE	EST.		
ORGANISM	soybean. glycine max Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.		
REFERENCE	1 (bases 1 to 547)		
AUTHORS	Shoemaker,R., Keim,P., Vodkin,L., Eipelting,J., Corryell,V., Khanna,A., Bolla,B., Maritz,M., Hillier,L., Kuababa,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,B., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harey,N., Schurr,R., Rittner,E., Kohn,S., Ship,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.		
TITLE	Public Soybean EST Project		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Possible reversed clone: similarity on wrong strand This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccuresgen.com High quality sequence stop: 423. Location/Qualifiers 1..547 /organism="Glycine max" /db_xref="taxon:3847" /clone="GENOME SYSTEMS CLONE ID: Gm-cl068-4294" /clone_lib="Gm-cl068" /tissue_type="Leaf, drought stressed, 1 month old plants,		
FEATURES			
SOURCE			

greenhouse grown"
/lab_host="DH10B"
/note="vector: pBluescript II SK+, Site_1: EcoRI, Site_2:
XhoI; The cDNA library was constructed from mRNA isolated
from drought stressed leaf tissue of the cultivar Williams
82. The month old greenhouse grown plants were deprived of
water for 3 days prior to harvesting the stressed leaf
tissue. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site. EcoRI adapters were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion. The
cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pBluescript vector. The
ligated cDNA fragments were transformed into DH10B host
cells (GibcoBRL). This library was constructed in the
laboratory of Dr. Randy Shoemaker."

BASE COUNT 123 a 141 c 133 g 150 t
ORIGIN

Query Match 17.2% Score 309.4; DB 10; Length 547;
Best Local Similarity 74.4%; Pred. No. 3.2e-81;
Matches 404; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

QY 991 aatcacccgagctcgagatcagcttgcagccagcagatgatgcagattcgcgatt 1050
||||| ||||||| ||||| ||||| ||||||| ||||||| ||||||| ||||||| ||
DB 541 AATATCCCCGAGCTGGAATTCCTTGAAAGAACCCACCATGATGACGATTCCGCATC 482
QY 1051 aagaggaacttcgttcgttcagcagcaagtcgcttcaagctccggaagatgac 1110
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 481 ACCGAGAACTTGTGTCCTCCCGACGAGGTGCTTCAAGTATACAGAAATGATC 422
QY 1111 cgcgctggctccggttggttgcagcaagaagtcgcgaagtcggaatttagac 1170
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 421 ACCGGAGGCTCCCGCTTCTACGACAAAGAACAGTTTCCAGATTCCGATTCGCAC 362
QY 1171 aatacgcgcgaagatcagcagacatlaagtgcagtcgcagattcgtctgc 1230
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 361 AAGATATGCTAAGATGCAATATGTAATGATGATGATGCCAGTCTTGTTTC 302
QY 1231 catctctggaagcgttcgggaagaagcagaagatgagtcgttcgtaaggttcct 1290
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 301 CACCTCTGGAAGCGCTGGAGAGACCGGAAATGATGATGATGCTCCTGCTCTCC 242
QY 1291 atgactccacccagactcaatttcacagagtcgagagaaatcgaagatgctctgt 1350
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 241 ATGACCCCGCGGACTCCATTTTCAACGAATCGAAGAGATTGAGAGCATTTGTTCG 182
QY 1351 gaatccgccttgatctcaaaacgcgttgatcaactcgcgcgcagatctcccaagaa 1410
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 181 GAGATTAAGGCTGAATTTGAAGACAGGCAAGTCCACAGAAACCCATTTATCTC---G 125
QY 1411 gatcaacaagtcaacctcgaaagcagggatgctcaacagaacaatcgcgcgttaaac 1470
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 124 TCCGAAACAAGTCAACTTGGAAACCGGATGTCACAGAAACAGCTCGGAGAAAGAGC 65
QY 1471 aactcgccttacttgcttgcagccggttcgaagctcaggaattcaggaattc 1530
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 64 AAGTTCGGGTACTTACACTTCCGAGCCGCTGTTTAAGGTTTGGGTTTCCCAAGTT 5
QY 1531 gat 1533
|||||
DB 4 GAT 2

RESULT 14
BE437072 BE437072 547 bp mRNA linear EST 18-MAY-2001
LOCUS EST408190 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
DEFINITION clone CLEB35K8, mRNA sequence.
ACCESSION BE437072
VERSION BE437072.1 GI:9434915
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; easterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 547)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Romling,C.M.,
Nierman,W., Frazer,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
Location/Qualifiers
source 1..547
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEB35K8"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="vector: pBluescriptSMCadapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the blossom end of
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

BASE COUNT 140 a 133 c 120 g 154 t
ORIGIN

Query Match 17.1% Score 308.6; DB 10; Length 547;
Best Local Similarity 72.8%; Pred. No. 5.6e-81;
Matches 398; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 431 aactcccgatccatcaaaaggagtgatgtgcgaacggaagcagcactccagc 490
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 1 AATATCCCAATATGTTCAAGCGCTTACGTTGCAAGCGACCTAACCTCTTTTGAAC 60
QY 491 cgggtacgaagtcacacactcttcgacggaagcagatagttcagcgcgtcaaatcgaac 550
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 61 CAACCGCCGACACATTTCTTCGACGGGACGATAGGTTCCGCCGTTCAATTCAAA 120
QY 551 acggttaagctagcagcttcggttcagctcaactcaacacggttttcaggaagc 610
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 121 ATGGGTCGGCTAGTTACGCTTGCCTTTCACAGAAACAGAGGCTTGTCAAGAAAG 180
QY 611 aatgtgcagcaggttttccccaagccatcgatgcgttcagcagcagcagcagatg 670
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 181 CTTTGGGTGGCCCTGTTTCCCTTAAGCCATTTGGATTTACATGTCACCTCGAATTG 240
QY 671 cccgacatgctatcttaacgacagagctgcagccggtatagtcgaccgagcagaa 730
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 241 CAAGGCTTATGCTGTTTACGCTCGTGGCTCTTGGACTTGTGATACACAGTAAGAA 300
QY 731 ccggtgtacgaacccggttttggtctatcttaatgagccggttatgtctatgctgag 790
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 301 CTGGGTGTCAAACGCCGGTTTATGCTTATTTCAATTAACCGATTACTGCTATGCTGAAG 360
QY 791 atgattacctaacaagtgcagatccactcccaatgcagattaaacccggtgtcgt 850
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 361 ATGATTTGCTTACCATGTGAAGGTAAACACCCACCGGAGATTTAAACAGAGGTCAT 420
QY 851 tcgattttgagacaattagaatcacaatgattgcaccacccgaagtgcagccgaat 910
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB 421 TCGATTTGACGCGCAGCTAAATATCACCATGATAGCTACCCAAAGCTCGACCAAGTTT 480

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 24, 2002, 03:54:00 ; Search time 57.66 Seconds
(without alignments)
7668.059 Million cell updates/sec

Title: US-09-758-269-5

Perfect score: 1800
Sequence: 1 atgcgtcttcacgcgaac.....tgcggaagcagctcgtgtga 1800

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCrUS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.8	2.2	7218	1 US-08-232-463-14	Sequence 14, Appl
2	34.8	1.9	4695	6 5225348-3	Patent No. 5225348
3	34.2	1.9	1221	1 US-08-434-881-1	Sequence 1, Appl
4	34.2	1.9	1221	3 US-08-977-771-1	Sequence 1, Appl
5	34.2	1.9	1221	4 US-09-361-773-1	Sequence 1, Appl
6	34.2	1.9	1370	4 US-08-026-408-12	Sequence 12, Appl
7	34.2	1.9	1371	4 US-09-026-408-1	Sequence 1, Appl
8	33.4	1.9	788	4 US-08-991-789A-177	Sequence 177, App
9	33.4	1.9	788	4 US-09-062-451-177	Sequence 177, App
10	33.4	1.9	1047	2 US-08-494-907-11	Sequence 11, Appl
11	33.4	1.9	1047	5 PCT-US96-10986-11	Sequence 11, Appl
12	33.4	1.9	1753	6 5225348-2	Patent No. 5225348
13	33.4	1.9	3680	2 US-08-494-907-1	Sequence 1, Appl
14	33.4	1.9	3680	5 PCT-US96-10986-1	Sequence 1, Appl
15	33.4	1.9	5076	2 US-08-494-907-2	Sequence 2, Appl
16	33.4	1.9	5076	5 PCT-US96-10986-2	Sequence 2, Appl
17	33.4	1.9	6170	2 US-08-494-907-4	Sequence 4, Appl
18	33.4	1.9	6170	5 PCT-US96-10986-4	Sequence 4, Appl
19	33.4	1.9	6387	2 US-08-494-907-3	Sequence 3, Appl
20	33.4	1.9	6387	5 PCT-US96-10986-3	Sequence 3, Appl
21	33.4	1.9	7198	4 US-08-994-035C-4	Sequence 4, Appl
22	33.4	1.9	4411529	4 US-09-103-840A-1	Sequence 1, Appl
23	33	1.8	607	4 US-09-385-982-525	Sequence 525, App
24	33	1.8	2128	2 US-08-371-377-15	Sequence 16, Appl
25	33	1.8	80161	3 US-09-036-967A-1	Sequence 1, Appl
26	33	1.8	80161	4 US-09-370-700-1	Sequence 1, Appl
27	32.6	1.8	1092	4 US-09-077-675A-15	Sequence 15, Appl

28	32.6	1.8	3129	4 US-09-077-675A-14	Sequence 14, Appl
29	31.8	1.8	7218	1 US-08-232-463-14	Sequence 14, Appl
30	31.6	1.8	1876	3 US-08-714-918-33	Sequence 33, Appl
31	31.6	1.8	1876	4 US-09-265-315-33	Sequence 33, Appl
32	31.6	1.8	1876	4 US-09-265-315-33	Sequence 33, Appl
33	31.6	1.8	1876	4 US-09-266-417-33	Sequence 33, Appl
34	31.4	1.7	30001	1 US-08-125-468-1	Sequence 1, Appl
35	31.4	1.7	30001	2 US-08-474-933-1	Sequence 1, Appl
36	31.2	1.7	3048	2 US-08-680-376-25	Sequence 25, Appl
37	31.2	1.7	3077	1 US-08-074-121-1	Sequence 1, Appl
38	31.2	1.7	3077	5 PCT-US94-06447-1	Sequence 1, Appl
39	31	1.7	289	4 US-09-007-005-17	Sequence 17, Appl
40	31	1.7	289	4 US-09-244-796-17	Sequence 17, Appl
41	30.4	1.7	551	4 US-09-328-111-448	Sequence 448, Appl
42	30.4	1.7	4316	1 US-08-317-450B-14	Sequence 14, Appl
43	30.4	1.7	4316	3 US-08-800-593-14	Sequence 14, Appl
44	30.4	1.7	5200	1 US-08-317-450B-12	Sequence 12, Appl
45	30.4	1.7	5200	3 US-08-800-593-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOAMPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpc-F15
US-08-232-463-14

```
Query Match          2.2%; Score 38.8; DB 1; Length 7218;
Best Local Similarity 7.9%; Pred. No. 0.049;
Matches 16; Conservative 112; Mismatches 74; Indels 0; Gaps 0;

OY 21 gctgcggttcctcgagatgctgtgccaatcactcagccgcattatcgtcttc 80
    ||||| | ||||| | | : : : : : : : : : : : : : : : : : : :
Db 1041 GCGTCGACAGTCGAGGAGCTTCGATYYYYYYYYYYYYYYYYYYYYYYYY 1100

OY 81 tcaagcgccagctgagttatttgccttaccctatgscagtcgtgacacgtaa 140
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1101 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1160

OY 141 gctcaatgttcacatcgctcgcacacccctccagctcttcattccctaaagcaatc 200
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1161 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1220

OY 201 aaactcccgccattgtgtc 222
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1221 YYYYYYYYYYYYYYYYYYYY 1242

RESULT 2
5225348-3
; Patent No. 5225348
; APPLICANT: HAGATA, SHIGERAZU; SUGANO, SUMIO; KIM, DONG W.;
; UETSUKI, TAICHI; KAZIRO, YOSHITO
; TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
; CONTAINING THE DNA FRAGMENT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,823
; FILING DATE: 08-DEC-1989
; SEQ ID NO: 3:
; LENGTH: 4695
5225348-3

Query Match          1.9%; Score 34.8; DB 6; Length 4695;
Best Local Similarity 48.1%; Pred. No. 0.78;
Matches 99; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

OY 340 cctaaacgctgatactcctagtttcaagatcgcgcggaatttgcctcggtagaataacag 399
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3386 cctgattgattgcacacagcgcctcattgcatgaagtttcgagctgaagaagaag 3445

OY 400 cccgcgcggcgaatctccggtgtgcgaataactcccgatccatcaagaagtgtat 459
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3446 attgacgcgcttcgtgtaaaaaagctggaagatgcccctaaattcttgaagtcgtgat 3505

OY 460 gtgcgcaacgagcctaaccactcaacgagccggtgacaggtcaccaactcttcgaagaa 519
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3506 gctgccattgtgatatagttcttcgcaagcccatggtgtgtagaagcttccagactat 3565

OY 520 gacggatggttacgcgcgtcaaat 545
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3566 ccaccttgggaagatgactactt 3591

RESULT 3
US-08-434-881-1/c
; Sequence 1, Application US/08434881
; Patent No. 5804376
; GENERAL INFORMATION:
; APPLICANT: Braxton, Scott M.
; APPLICANT: Wille, Craig G.
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Pancreas-Derived Serpin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3330 Hillview Avenue
; CITY: Palo Alto
; STATE: California
```

```
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/434,881
FILING DATE: Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Luther, Barbara J.
REGISTRATION NUMBER: 33954
REFERENCE/DOCKET NUMBER: PF0035 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-852-0195
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1221 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
LIBRARY: Pancreas
CLONE: 222689
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1221
US-08-434-881-1

Query Match          1.9%; Score 34.2; DB 1; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 97 agttttgagtccttaccatagtcgccaatgctgttcacagtaagctcaatgttcaatct 156
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 788 ACTTCTTCTATATCCATACCTTCTGCAGAAAGTATGATATTAAGCTAATTCATCACCT 729

OY 157 gcgcctacactccctcagctcttcattccctaaagcaatcaactcccgccatt 216
    | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 TTGTAAGACAATTCTAAACTGTGATTCAGAGGAAGATTCAGAATAATTAACATATT 669

OY 217 gttgtta 223
    | | | | |
Db 668 GTTCTCA 662

RESULT 4
US-08-977-771-1/c
; Sequence 1, Application US/08977771
; Patent No. 6013448
; GENERAL INFORMATION:
; APPLICANT: Braxton, Scott M.
; APPLICANT: Wille, Craig G.
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Pancreas-Derived Serpin
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3330 Hillview Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
```

```

:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/977,771
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/434,881
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Luther, Barbara J.
: REGISTRATION NUMBER: 33954
: REFERENCE/DOCKET NUMBER: PF0035 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-852-0195
: TELEFAX: 415-852-0195
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1221 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: IMMEDIATE SOURCE:
: LIBRARY: Pancreas
: CLONE: 222689
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1221
:
: US-08-977-771-1

```

```

Query Match          1.9%; Score 34.2; DB 3; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 97 agtatttgagctccttaccatagcgagtcgtgcacagcgaagctcaatgttcatct 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 788 ACTCTTCTATATCATCTCTCTGCGAGAGTATGATTAATTAATTAATTCATCACCCT 729

QY 157 ggccttcacacccctccagctcttcattcccttaagcaatcatcaactcccgccatt 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 728 TTGTAAAGCAATTTAAACCTTGTAAGTTTCAGGAGATTCAGAAAATAACCATATTTT 669

QY 217 gttgtta 223
    |||||
DB 668 GTTCTCA 662

```

```

RESULT 5
US-09-361-773-1/c
: Sequence 1, Application US/09361773
: Patent No. 6197519
: GENERAL INFORMATION:
: APPLICANT: Braxton, Scott M.
: APPLICANT: Wilde, Craig G.
: APPLICANT: Diep, Dinh
: TITLE OF INVENTION: Pancreas-Derived Serpin
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3330 Hillview Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PC-DOS/MS-DOS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/361,773
: FILING DATE:
: CLASSIFICATION:

```

```

:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/977,771
: FILING DATE:
: CLASSIFICATION:
: APPLICATION NUMBER: 08/434,881
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Luther, Barbara J.
: REGISTRATION NUMBER: 33954
: REFERENCE/DOCKET NUMBER: PF0035 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-852-0195
: TELEFAX: 415-852-0195
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1221 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: IMMEDIATE SOURCE:
: LIBRARY: Pancreas
: CLONE: 222689
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1221
:
: US-09-361-773-1

```

```

Query Match          1.9%; Score 34.2; DB 4; Length 1221;
Best Local Similarity 54.3%; Pred. No. 0.51;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 97 agtatttgagctccttaccatagcgagtcgtgcacagcgaagctcaatgttcatct 156
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 788 ACTCTTCTATATCATCTCTCTGCGAGAGTATGATTAATTAATTAATTCATCACCCT 729

QY 157 ggccttcacacccctccagctcttcattcccttaagcaatcatcaactcccgccatt 216
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 728 TTGTAAAGCAATTTAAACCTTGTAAGTTTCAGGAGATTCAGAAAATAACCATATTTT 669

QY 217 gttgtta 223
    |||||
DB 668 GTTCTCA 662

```

```

RESULT 6
US-09-026-408-12/c
: Sequence 12, Application US/09026408
: Patent No. 6303338
: GENERAL INFORMATION:
: APPLICANT: Ni et al.
: TITLE OF INVENTION: PANCREAS-DERIVED PLASMINOGEN ACTIVATOR
: TITLE OF INVENTION: INHIBITOR
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
: STREET: 1100 NEW YORK AVENUE, SUITE 600
: CITY: WASHINGTON
: STATE: DC
: COUNTRY: USA
: ZIP: 20005-3934
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/026,408
: FILING DATE: Herewith
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/934,011
: FILING DATE: 15-AUG-1997

```



```

; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
;   NAME: Potter, Jane E. R.
;   REGISTRATION NUMBER: 33,332
;   REFERENCE/DOCKET NUMBER: 210121.419C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 682-4900
;   TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 177:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 788 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 177:
US-08-991-789A-177

Query Match      1.9%; Score 33.4; DB 4; Length 788;
Best Local Similarity 48.2%; Pred. No. 0.71;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 340 cctaaacgctgatacctagtgctcagatgcgcggaatttgcgcggtgaatgaacag 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 CCTGATTGATGTCACACACGCTCACAATTGCATTCAGATTGCTGAGCTGAAGGAAAAG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 cccgctcggcgttaatcttcgcggtgctcgaaacttcccgatccatcaagaagtgat 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 ATTGATGCGCCGTTCTGTGTAATAAAGCTGGAAGATGGCCCTTAATTCTTAAATCTGCTGAT 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 gtgcgaacgagagctaacccactcaacgagccggtgacaggtcaccacttctcgacgga 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GCTGCCATTGTTGATATGTTCTGTGCAAGCCCATGTGTGAGAGCTTCTCAGACTAT 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 gacggtatggttcac 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 CCACCTTTGGGTGCG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-09-062-451-177
; Sequence 177, Application US/09062451
; Patent No. 6344550
; GENERAL INFORMATION:
;   APPLICANT: Frudakis, Tony N.
;   APPLICANT: Smith, John M.
;   APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;   TREATMENT OF INFECTIONS: TREATMENT AND DIAGNOSIS OF BREAST CANCER
; NUMBER OF SEQUENCES: 297
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: SEED and BERRY LLP
;   STREET: 6300 Columbia Center, 701 Fifth Avenue
;   CITY: Seattle
;   STATE: Washington
;   COUNTRY: USA
;   ZIP: 98104-7092
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/062.451
;   FILING DATE: 04-Apr-1997
; CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Makl, David J.
;     REGISTRATION NUMBER: 31,392
;     REFERENCE/DOCKET NUMBER: 210121.419C2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (206) 622-4500
;   TELEFAX: (206) 682-6031
```

```

; INFORMATION FOR SEQ ID NO: 177:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 788 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; US-09-062-451-177

Query Match      1.9%; Score 33.4; DB 4; Length 788;
Best Local Similarity 48.2%; Pred. No. 0.71;
Matches 94; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 340 cctaaacgctgatacctagtgctcagatgcgcggaatttgcgcggtgaatgaacag 399
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 536 CCTGATTGATGTCACACACGCTCACAATTGCATTCAGATTGCTGAGCTGAAGGAAAAG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 400 cccgctcggcgttaatcttcgcggtgctcgaaacttcccgatccatcaagaagtgat 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 ATTGATGCGCCGTTCTGTGTAATAAAGCTGGAAGATGGCCCTTAATTCTTAAATCTGCTGAT 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 gtgcgaacgagagctaacccactcaacgagccggtgacaggtcaccacttctcgacgga 519
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GCTGCCATTGTTGATATGTTCTGTGCAAGCCCATGTGTGAGAGCTTCTCAGACTAT 715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 gacggtatggttcac 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 CCACCTTTGGGTGCG 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-494-907-11/c
; Sequence 11, Application US/08494907
; Patent No. 5955298
; GENERAL INFORMATION:
;   APPLICANT: Thomasow, Linda S
;   APPLICANT: Banger, Mahalaxmi
;   APPLICANT: Weller, David M
;   APPLICANT: Cook, R. James
; TITLE OF INVENTION: Sequences for Production of
;   2,4-Diacetylphloroglucinol and Methods
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Margaret A. Connor, USDA-ARS
;   STREET: 800 Buchanan Street
;   CITY: Albany
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94710
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/494.907
;   FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Connor, Margaret A
;   REGISTRATION NUMBER: 30043
;   REFERENCE/DOCKET NUMBER: 0009.95
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 559-6067
;   TELEFAX: (510) 559-5777
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1047 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
```



```

1  APPLICANT:  Weller, David M
2  APPLICANT:  Cook, R. James
3  TITLE OF INVENTION:  Sequences for Production of
4  TITLE OF INVENTION:  2,4-Diacetylphloroglucinol and Methods
5  NUMBER OF SEQUENCES:  20
6  CORRESPONDENCE ADDRESSES:
7  ADDRESSEE:  Margaret A. Connor, USDA-ARS
8  STREET:  800 Buchanan Street
9  CITY:  Albany
10 STATE:  CA
11 COUNTRY:  USA
12 ZIP:  94710
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  PatentIn Release #1.0, Version #1.25
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/494,907
20 FILING DATE:
21 CLASSIFICATION:  435
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Connor, Margaret A
24 REGISTRATION NUMBER:  30043
25 REFERENCE/DOCKET NUMBER:  0009.95
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  (510) 559-6067
28 TELEFAX:  (510) 559-5777
29 INFORMATION FOR SEQ ID NO:  1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  3680 base pairs
32 TYPE:  nucleic acid
33 STRANDEDNESS:  single
34 TOPOLOGY:  linear
35 MOLECULE TYPE:  DNA (genomic)
36 HYPOTHETICAL:  NO
37 ANTI-SENSE:  NO
38 ORIGINAL SOURCE:
39 ORGANISM:  Pseudomonas fluorescens
40 FEATURE:
41 NAME/KEY:  misc_feature
42 LOCATION:  complement(2889..3680)
43 OTHER INFORMATION:  /note="ph1a, transcribed from
44 OTHER INFORMATION:  right to left"
45 FEATURE:
46 NAME/KEY:  misc_feature
47 LOCATION:  complement(1210..1917)
48 OTHER INFORMATION:  /note="ph1b, transcribed from
49 OTHER INFORMATION:  right to left"
50 FEATURE:
51 NAME/KEY:  misc_feature
52 LOCATION:  complement(1689..2855)
53 OTHER INFORMATION:  /note="ph1c, transcribed from
54 OTHER INFORMATION:  right to left"
55 FEATURE:
56 NAME/KEY:  misc_feature
57 LOCATION:  2118..3371
58 OTHER INFORMATION:  /note="ph1r, transcribed from left
59 OTHER INFORMATION:  to right"
60 FEATURE:
61 NAME/KEY:  misc_feature
62 LOCATION:  1..3680
63 OTHER INFORMATION:  /note="SEQ ID NO:1 contains genes
64 OTHER INFORMATION:  necessary for Ph1 synthesis."
65 US-08-494-907-1

```

	Best Local Similarity	48.2%	Pred. No. 1.9;	Matches 94;	Conservative 0;	Mismatches 101;	Indels 0;	Gaps 0;
QY	585	gactaacccggtttgttcgagacgtcaatttggtgttcgaccggttttccccaagcatcg	644					
Db	938	GGCTAAACGGATGATCTAAACACCCAGATCANTAAACGATATCGTTTCCCATGCA	879					
QY	645	tgaagcttaacgccaacacccgattatgtcccgactcatctatctaacccaagcctgcagc	704					
Db	878	TGAACCTTCGGGTGCATACCGGCTTAAACCACCGCAGCATCGTGTATGACGAGAGGCTCG	819					
QY	705	cgtatagtcgacccgcacacagcagccggtgtagctaacgcggtgtgtctattca	764					
Db	818	CCGCATGCTTCAATCGCCCGCGCGCCAGGCCATTAAAAACGGGGCTTACACACGAGACA	759					
QY	765	tgagccggtattatgc	779					
Db	758	CATCCGGATGGTCGC	744					

```

RESULT 14
PCT-US96-10986-1/c
: Sequence 1, Application PC/TUS9610986
: GENERAL INFORMATION:
: TITLE OF INVENTION: Sequences for Production of
: TITLE OF INVENTION: 2,4-Diacetylphloroglucinol and Methods
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Stephan A. Pendorf, DOMINIK & STEIN
: STREET: 600 N. West Shore Boulevard, Suite 1000
: CITY: Tampa
: STATE: FL
: COUNTRY: USA
: ZIP: 33609
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/10986
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Pendorf, Stephan A.
: REGISTRATION NUMBER: 32665
: REFERENCE/DOCKET NUMBER: A700.320
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (813) 289-2966
: TELEFAX: (813) 289-2966
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3680 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Pseudomonas fluorescens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: complement (2889..3680)
: OTHER INFORMATION: /note="phb, transcribed from
: OTHER INFORMATION: right to left"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: complement (1210..1917)
: OTHER INFORMATION: /note="phb, transcribed from
: OTHER INFORMATION: right to left"
: FEATURE:
: NAME/KEY: misc_feature

```

REGISTRATION NUMBER: 30043

Db 2154 CATCCGGATGGTCGC 2140

Search completed: July 24, 2002, 06:21:27
Job time: 8847 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:11:48 ; Search time 34.28 Seconds

(without alignments)
3022.869 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150

Sequence: 1 MASFTATAVSGRWLGSGNHT.....VPYGHGTFIGADDLAKQV 599

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mnc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3150	100.0	599	10	Q9LR77 arabidopsis
2	3143	99.8	599	10	Q93ZU5 arabidopsis
3	2289	72.7	604	10	Q9M3Z9 solanum tub
4	2280.5	72.4	605	10	Q240Z3 lycopersico
5	2170.5	68.9	657	10	Q9M9F5 arabidopsis
6	2168.5	68.8	612	10	Q9F5Z4 arabidopsis
7	2165	68.7	615	10	Q9M6E8 vigna unguis
8	2066.5	65.6	625	10	Q9M6E8 phaseolus v
9	2044.5	64.9	589	10	Q9C6Z1 persea amer
10	1991	63.2	593	10	Q9C6Z1 arabidopsis
11	1977	62.8	589	10	Q9AXX3 persea amer
12	1930	61.3	604	10	Q24592 zea mays (m
13	1663.5	52.8	577	10	Q9LRM7 arabidopsis
14	968	30.7	595	10	Q49675 arabidopsis
15	959.5	30.5	543	10	Q94IR2 phaseolus v
16	938	29.8	538	10	Q9LY63 arabidopsis

17	937	29.7	538	10	Q65572 arabidopsis
18	837.5	26.6	200	10	Q94EN8 lactuca sat
19	798.5	25.3	524	10	Q9AXZ5 persea amer
20	745	23.7	446	10	Q49895 malus domes
21	528.5	16.8	483	16	Q9AA32 caulobacter
22	527.5	16.7	456	2	Q9AM11 streptomyces
23	526.5	16.7	456	2	Q93FA4 streptomyces
24	466	14.8	501	16	Q06785 mycobacteri
25	464	14.7	503	2	Q9RK46 streptomyces
26	454.5	14.4	502	16	Q05905 mycobacteri
27	394	12.5	490	16	P74334 synecocyst
28	352	11.2	485	2	Q53353 pseudomonas
29	317	10.1	480	16	P74370 synecocyst
30	303.5	9.6	490	2	Q52008 pseudomonas
31	302.5	9.6	552	10	Q93VDS oryza sativ
32	270	8.6	542	2	Q68868 synecococc
33	260	8.3	526	13	Q91993 gallus galli
34	256	8.1	516	13	Q90WH4 brachydanio
35	253	8.0	616	10	Q9M079 arabidopsis
36	244	7.7	618	10	Q9SHD9 arabidopsis
37	238.5	7.6	549	13	Q90WH3 brachydanio
38	238.5	7.6	556	5	Q9TX79 caenorhabditis
39	232.5	7.4	532	11	Q9NFI1 mus musculus
40	231	7.3	483	16	Q9AB98 caulobacter
41	225	7.1	547	4	Q9AY66 mus musculus
42	224	7.1	566	11	Q9TJ56 mus musculus
43	223	7.0	547	4	Q9NVH5 mus musculus
44	221	7.0	566	11	Q9ERN9 mus musculus
45	216	6.9	533	13	Q9Y125 ambystoma t

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	599 AA.
Q9LR77	Q9LR77	01-OCT-2000 (TRENBLREL. 15, Created)		
Q9LR77	Q9LR77	01-OCT-2000 (TRENBLREL. 15, Last sequence update)		
Q9LR77	Q9LR77	01-DEC-2001 (TRENBLREL. 19, Last annotation update)		
Q9LR77	Q9LR77	9-CIS-EPOXYCAROTENOID DIOXYGENASE (NEOXANTHIN CLEAVAGE ENZYME).		
Q9LR77	Q9LR77	Arabidopsis thaliana (Mouse-ear cress).		
Q9LR77	Q9LR77	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:		
Q9LR77	Q9LR77	OC: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:		
Q9LR77	Q9LR77	OC: eucotids II: Brassicales: Brassicaceae: Arabidopsis.		
Q9LR77	Q9LR77	NCBI_TaxID=3702;		
Q9LR77	Q9LR77	[1]		
Q9LR77	Q9LR77	SEQUENCE FROM N.A.		
Q9LR77	Q9LR77	RC STRAIN-COLUMBIA:		
Q9LR77	Q9LR77	Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.:		
Q9LR77	Q9LR77	Submitted (Jun-1999) to the EMBL/GenBank/DBJ databases.		
Q9LR77	Q9LR77	[2]		
Q9LR77	Q9LR77	SEQUENCE FROM N.A.		
Q9LR77	Q9LR77	RC STRAIN-COLUMBIA:		
Q9LR77	Q9LR77	Medline=20277480; PubMed=10819329;		
Q9LR77	Q9LR77	Nakamura Y.:		
Q9LR77	Q9LR77	"Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence		
Q9LR77	Q9LR77	features of the regions of 4,504,864 bp covered by sixty p1 and TAC		
Q9LR77	Q9LR77	clones."		
Q9LR77	Q9LR77	DNA Res. 7.131-135(2000).		
Q9LR77	Q9LR77	[3]		
Q9LR77	Q9LR77	SEQUENCE FROM N.A.		
Q9LR77	Q9LR77	RC STRAIN-COLUMBIA:		
Q9LR77	Q9LR77	Iuchi S., Kobayashi M., Shinozaki K.:		
Q9LR77	Q9LR77	"Characterization of neoxanthin cleavage enzyme from Arabidopsis		
Q9LR77	Q9LR77	thaliana."		
Q9LR77	Q9LR77	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.		
Q9LR77	Q9LR77	EMBL: AB026817; BAB01336.1; -		
Q9LR77	Q9LR77	EMBL: AB026549; BAB70609.1; -		
Q9LR77	Q9LR77	InterPro: IPR004294; RPE65.		
Q9LR77	Q9LR77	Pfam: PF03055; RPE65; 1.		
Q9LR77	Q9LR77	Dioxygenase.		

SQ SEQUENCE 599 AA: 65856 MW: 7D513E39945E0CF3 CRC64:

Query Match 100.0%; Score 3150; DB 10; Length 599;
Best Local Similarity 100.0%; Pred. No. 1.7e-242;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
DB 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
QY 61 HEPKOSNSPATAIVYKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKHLHPKPTADS 120
DB 61 HEPKOSNSPATAIVYKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKHLHPKPTADS 120
QY 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGHFFDGDGVHAY 180
DB 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGHFFDGDGVHAY 180
QY 181 KEHGSASACRPTQTNRFVQERQLGRPVFPKAIIGELHGTGIALRLMIFYARAAAGIVDP 240
DB 181 KEHGSASACRPTQTNRFVQERQLGRPVFPKAIIGELHGTGIALRLMIFYARAAAGIVDP 240
QY 241 AHGTGVANAGLYFNGRLAMSDDLRYOQVITPNNGDLTVGFRFDPDGOLESTMIHPKV 300
DB 241 AHGTGVANAGLYFNGRLAMSDDLRYOQVITPNNGDLTVGFRFDPDGOLESTMIHPKV 300
QY 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
QY 361 QVYFKLPEMIRGSPVYVYKNNKVARFGIILDKYAEDSSNKKWIDAPCCFCFHLNAMEEPE 420
DB 361 QVYFKLPEMIRGSPVYVYKNNKVARFGIILDKYAEDSSNKKWIDAPCCFCFHLNAMEEPE 420
QY 421 TDEVVVIGSCMPDPSIFNESDENLKSYLEIRLNKTEGSTRRLPIISNEDQVNLKAGM 480
DB 421 TDEVVVIGSCMPDPSIFNESDENLKSYLEIRLNKTEGSTRRLPIISNEDQVNLKAGM 480
QY 481 VNRNMLGRRTKFAVYLLALAPWPKVSGFAKVDLTGCVKKNHLYGDNRYGGEPLFLPGEGE 540
DB 481 VNRNMLGRRTKFAVYLLALAPWPKVSGFAKVDLTGCVKKNHLYGDNRYGGEPLFLPGEGE 540
QY 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEAVTVKLPSPVPYGFHGTFTGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEAVTVKLPSPVPYGFHGTFTGADDLAKQV 599

RESULT 2
QY 093205 PRELIMINARY; PRT: 599 AA.
AC 093205;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE PUTATIVE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
GN MOA2.4/AT3G14440.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A.
RA Yamada K., Banb J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk K., Hayashizaki Y.,
RA Ishida J., Jones T., Kaniya A., Karlin-Neumann G., Kawai J., Kim C.,
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,
RA Nguyen M., Pam C.J., Sakurai T., Satou M., Seki M., Shin P.,
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.,
RT "Full Length cDNA of gene MOA2.4/AT3G14440 (GI:11994214).";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056255; AAL07104.1; -
KM Dioxxygenase.
SQ SEQUENCE 599 AA: 65813 MW: B9007A2DC1C15506 CRC64:

Query Match 99.8%; Score 3143; DB 10; Length 599;
Best Local Similarity 99.8%; Pred. No. 6.3e-242;
Matches 598; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
DB 1 MASFTATAAVSGRWLGNGHTQPPPLSSQSSDLSYCSLPMASHVTRKLVNSALHTPPAL 60
QY 61 HEPKOSNSPATAIVYKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKHLHPKPTADS 120
DB 61 HEPKOSNSPATAIVYKKRAKESNTKQNNLQRAAAALDAAEGFLVSHKHLHPKPTADS 120
QY 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGHFFDGDGVHAY 180
DB 121 VOIAGNFAVNEQPVRRNLPPVVGKLPDSIKGVYVRNGANPLHEPVYGHFFDGDGVHAY 180
QY 181 KEHGSASACRPTQTNRFVQERQLGRPVFPKAIIGELHGTGIALRLMIFYARAAAGIVDP 240
DB 181 KEHGSASACRPTQTNRFVQERQLGRPVFPKAIIGELHGTGIALRLMIFYARAAAGIVDP 240
QY 241 AHGTGVANAGLYFNGRLAMSDDLRYOQVITPNNGDLTVGFRFDPDGOLESTMIHPKV 300
DB 241 AHGTGVANAGLYFNGRLAMSDDLRYOQVITPNNGDLTVGFRFDPDGOLESTMIHPKV 300
QY 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
DB 301 DPESGELFALSYDVVSKPYLKTYFRSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 360
QY 361 QVYFKLPEMIRGSPVYVYKNNKVARFGIILDKYAEDSSNKKWIDAPCCFCFHLNAMEEPE 420
DB 361 QVYFKLPEMIRGSPVYVYKNNKVARFGIILDKYAEDSSNKKWIDAPCCFCFHLNAMEEPE 420
QY 421 TDEVVVIGSCMPDPSIFNESDENLKSYLEIRLNKTEGSTRRLPIISNEDQVNLKAGM 480
DB 421 TDEVVVIGSCMPDPSIFNESDENLKSYLEIRLNKTEGSTRRLPIISNEDQVNLKAGM 480
QY 481 VNRNMLGRRTKFAVYLLALAPWPKVSGFAKVDLTGCVKKNHLYGDNRYGGEPLFLPGEGE 540
DB 481 VNRNMLGRRTKFAVYLLALAPWPKVSGFAKVDLTGCVKKNHLYGDNRYGGEPLFLPGEGE 540
QY 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEAVTVKLPSPVPYGFHGTFTGADDLAKQV 599
DB 541 EDEGYILCFVHDEKTKWSELOIVNAVSLVEAVTVKLPSPVPYGFHGTFTGADDLAKQV 599

RESULT 3
QY 09M329 PRELIMINARY; PRT: 604 AA.
AC 09M329;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE PUTATIVE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
GN NCEB1.
OS Solanum tuberosum (potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasteride I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN SEQUENCE FROM N.A.
RA Burdige A., Taylor I.B., Thompson A.;
RT "Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ276244; CAB76920.1;
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.

Dioxygenase.
KW SEQUENCE 604 AA; 67288 MW; 36856BC82E1604A1 CRC64;

Query Match 72.7%; Score 2289; DB 10; Length 604;
Best Local Similarity 71.3%; Pred. No. 7.6e-174;
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;

```
QY 5 TATTAAGSGRMVGGNHTOPPLSSSSSDLSYCS---SLPMASVTRKLVSSALHTPPALH 61
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   3 TTSHATNTWTI-----KPKLSMPSSKKEGFASNSISLKNHNSQSLNINSLSQAPRLH 57
DB
QY 62 FPKOSSN---SPAIVKPKAKESNT---KOMLFPRAAALDAAGFLVSHKELHP 113
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   58 FPKOSSNQTPTKSTISHPKQENNSSSSISKWNLVQKAAMALDAVAGLTKHELHP 117
DB
QY 114 PKTADPSVQIAGNFPVNEQPVRRNLPVYGLKLPDSIKGVYVRNGANPLHEPVTHGHPFD 173
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   118 PKTADPSVQIAGNFPVNEQVSLPVTKIPKCVQGVYVRNGANPLHEPTAGRHHPFDG 177
DB
QY 174 DGMVAHAFEEHSGASVACRFPTQTRFVOEROLGRVPFPAKAGELGHGTIARLMIFYARA 233
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   178 DGMVAHAFKNGSASVACRFETERFVOEROLGRVPFPAKAGELGHGTIARLMIFYARG 237
DB
QY 234 AAGIVDPAGHGTGVANAGLVYFNGRLAMS EDDLPYQVQITPENGDLKTVGRFDPGQLEST 293
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   238 LFGILIDHSRGTVANAGLVYFNNRLAMSEDDLPHYHVKVPTFGDLKTEGRFDPGQLEST 297
DB
QY 294 MIAHPKVDPESELFALSYDVVSKPYLKYFRSPDGTSPVETIQLDQPTMHDFATIEK 353
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   298 MIAHPKLDVPSELFPALSYDVYQKPYLKYFRSKNGKSNVEIPEVDPMTMHDFATIEK 357
DB
QY 354 FVVVPDQGVNKLPEMIRGSGSPVYVDKKNVAFGLIDKYAEDSSNIKVIDAPDCFCFLM 413
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   358 FVILPDQGVNKLPEMIRGSGSPVYVDKKNVAFGLIDKYAEDSSNIKVIDAPDCFCFLM 417
DB
QY 414 NAMEBPEDEVVYVIGSCMTPPDSIFNESDENLKSVLSIRLNLTKGESTRRPIISNEDQ 473
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   418 NAMEBPEDEVVYVIGSCMTPPDSIFNECEDEGLKSVLSIRLNLTKGESTRRPIISNEDQ 477
DB
QY 474 VNLEAGVNNRMGLGRKTFAYLALAEPPPKVSGFAKVDLTGVEYKHLYGDNRYGGEPL 533
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   478 VNLEAGVNNRMGLGRKTFAYLALAEPPPKVSGFAKVDLTGVEYKHLYGDNRYGGEPL 537
DB
QY 534 LPGE--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVTLPSRVPGFGHGTIGA 591
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   538 LPROPNSKEEDDGYILAFVHDEKTKWSELQIVNAVSLVEATVTLPSRVPGFGHGTIGA 597
DB
QY 592 DDLAKQ 597
   1 1 1 1 1
DB 598 NDLANQ 603
```

RESULT 4
ID 024023 PRELIMINARY; PRT; 605 AA.
AC 024023;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NINE-CIS-EPOXYCAROTENOID DIOXYGENASE.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RA Burbidge A., Griewe T.M., Jackson A., Thompson A., Taylor I.B.;
RT "Structure and expression of a cDNA encoding a putative neoxanthin
cleavage enzyme (NCE) isolated from a wilt-related tomato
Lycopersicon esculentum Mill.) library.";
RL J. Exp. Bot. 47:2111-2112(1997).

[2]
RN SEQUENCE FROM N.A.
RP Burbidge A.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; 297215; CAB10168.1; -
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
KW Dioxygenase.

QY SEQUENCE 605 AA; 67317 MW; 19BA67930346A872 CRC64;

Query Match 72.4%; Score 2280.5; DB 10; Length 605;
Best Local Similarity 71.7%; Pred. No. 3.7e-173;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

```
QY 5 TATTAAGSGRMVGGNHTOPPLSSSSSDLSYCS---SLPMASVTRKLVSSALHTPPALH 61
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   3 TTSHATNTWTI-----KPKLSMPSSKKEGFASNSISLKNHNSQSLNINSLSQAPRLH 57
DB
QY 62 FPKOSSN--SPA--IVKPKAKESN---TKOMLFPRAAALDAAGFLVSHKELHP 112
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   58 FPKOSSNQTPTKNNITISHPKQENNSSSSSTKWNLVQKAAMALDAVAGLTKHELHP 117
DB
QY 113 LPKTADPSVQIAGNFPVNEQPVRRNLPVYGLKLPDSIKGVYVRNGANPLHEPVTHGHPFD 172
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   118 LPKTADPSVQIAGNFPVNEQVSLPVTKIPKCVQGVYVRNGANPLHEPTAGRHHPFD 177
DB
QY 173 DGMVAHAFEEHSGASVACRFPTQTRFVOEROLGRVPFPAKAGELGHGTIARLMIFYARA 232
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   178 DGMVAHAFKNGSASVACRFETERLVOEROLGRVPFPAKAGELGHGTIARLMIFYARA 237
DB
QY 233 AAGIVDPAGHGTGVANAGLVYFNGRLAMS EDDLPYQVQITPENGDLKTVGRFDPGQLES 292
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   238 GILFGLVDSKSGVANNAGLVYFNNRLAMS EDDLPHYHVKVPTFGDLKTEGRFDPGQLES 297
DB
QY 293 TWIAHPKVDPESELFALSYDVVSKPYLKYFRSPDGTSPVETIQLDQPTMHDFATIE 352
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   298 TWIAHPKLDVPSELFPALSYDVYQKPYLKYFRSKNGKSNVEIPEVDPMTMHDFATIE 357
DB
QY 353 NVVVPDQGVNKLPEMIRGSGSPVYVDKKNVAFGLIDKYAEDSSNIKVIDAPDCFCFL 412
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   358 NVVVPDQGVNKLPEMIRGSGSPVYVDKKNVAFGLIDKYAEDSSNIKVIDAPDCFCFL 417
DB
QY 413 NAMEBPEDEVVYVIGSCMTPPDSIFNESDENLKSVLSIRLNLTKGESTRRPIISNEDQ 472
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   418 NAMEBPEDEVVYVIGSCMTPPDSIFNECEDEGLKSVLSIRLNLTKGESTRRPIISNEDQ 477
DB
QY 473 QVNLEAGVNNRMGLGRKTFAYLALAEPPPKVSGFAKVDLTGVEYKHLYGDNRYGGEPL 532
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   478 QVNLEAGVNNRMGLGRKTFAYLALAEPPPKVSGFAKVDLTGVEYKHLYGDNRYGGEPL 537
DB
QY 533 FLPGE--GGEDEGYILCFVHDEKTKWSELQIVNAVSLVEATVTLPSRVPGFGHGTIGA 590
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
   538 FLPROPNSKEEDDGYILAFVHDEKTKWSELQIVNAVSLVEATVTLPSRVPGFGHGTIGA 597
DB
QY 591 ADDLAKQ 597
   1 1 1 1 1
DB 598 ANDLANQ 604
```

RESULT 5
ID 09M9F5 PRELIMINARY; PRT; 657 AA.
AC 09M9F5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE F3F9.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;

RA [1]
RA SEQUENCE FROM N.A.
RA Chao Q., Brooks S., Buehler E., Johnson-Hopson C., Khan S., Kim C.,
RA Shun P., Altafi H., Bel B., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA Ecker J.R.;
RA "Genomic sequence for Arabidopsis thaliana BAC F3p9 from chromosome
RT 1";
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RA Ecker J.R.;
RN Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RP
RP SEQUENCE FROM N.A.
RA Cheuk R., Shun P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzales A., Hansen N., Howng B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC013430; AAF11797.1; -
DR InterPro: IPR01899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING: UNKNOWN_1.
SQ SEQUENCE 657 AA; 73015 MW; F41DECBE94806318 CRC64;

Query Match 68.9%; Score 2170.5; DB 10; Length 657;
Best Local Similarity 67.7%; Pred. No. 2.4e-164;
Matches 419; Conservative 80; Mismatches 87; Indels 33; Gaps 9;

OY 1 MASFATAAAGSGRWL-----GNNTPPLSSQSSDLSYCSLP-----MASRYT 45
DB 51 MASTLLPSTSTQFLDRTSTSSSSSRPKLOS-----LSFSSTLRKKKLIVPCYVSSSVN 105
OY 46 RLANTSSAHT-----PPALHFPKSSNSPAIYVKKAKKSNTKOMLPORAAAALDAE 101
DB 106 KSSVSSSSQSSPTFKPPSR--KKLCNDYTNLI-PRTTNQNPK-LMPVORTAAMVDAVE 160
OY 102 GFLVSEH-KLHLPKTDADPSVOIAGNFAPVNEQPYARNLPVVGKLPDSIKGYVVRNGANP 160
DB 161 NAMISHERRRHHPKTDADPAVOIAGNFFRPVREKPYVHNLPVGTGYVECIQGYVVRNGANP 220
OY 161 LHEPYTGHFFPDGDMVAHVAFKFEHGSASACRFQTQNRFOERQUGRPVFPKAIQELGH 220
DB 221 LHKPVSGHLEFDGDMVAHVAFKFEHGSASACRFQTQNRFOERQUGRPVFPKAIQELGH 280
OY 221 TGIARLMFLYARAAGIYDPAHGTGVANAGLYFNGRLLAAMEDDLPRYOVTTPNGDLKT 280
DB 281 LGIARLMFLYARAAGIYDPAHGTGVANAGLYFNGRLLAAMEDDLPRYHVKTQYGDLETT 340
OY 281 VGRFPDGLSESTMIAPKRVDPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 340
DB 341 SGRYPDGLSKSTMIAPKRVDPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 400
OY 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKKNVAREGILDKYAEDSSNIK 400
DB 401 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKKNVAREGILDKYAEDSSNIK 460
OY 401 WIDADPCFCFHLNMAWEEETDEVVVIGSCMTPPSIFNESDENLKSYLEIRLNLKTGE 460
DB 461 WIDADPCFCFHLNMAWEEETDEVVVIGSCMTPPSIFNESDENLKSYLEIRLNLKTGE 520
OY 461 STRRPILSNEDQOVULEAGVNRNMLGRKTKPAPYALALAEPMKVSQGFPAKVDLTGGEVKKH 520
DB 521 STRRPILSNEDQOVULEAGVNRNMLGRKTKPAPYALALAEPMKVSQGFPAKVDLTGGEVKKH 577
OY 521 LYGNRRYGGEPFLPFGEGEEDGYILCFVHDEKTKWSELOIVNVSLEVEATVKLPSRV 580

DB 578 IYGEKYGGEPFLPFGSGDEEDGYIMVHDEEKVKSRLQILNANMNLKXATVTLPSRV 637
OY 581 PYGHGTFTGADDLAKQVY 599
DB 638 PYGHGTFTSKEDLSKQAL 656
RESULT 6
OYFS24 PRELIMINARY; PRT; 612 AA.
AC OYFS24;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEOXANTHIN CLEAVAGE ENZYME.
GN CPRD65.
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20317197; PubMed=10859185;
RA Iuchi S., Kobayashi M., Yamaguchi-Shinozaki K., Shinozaki K.;
RT "A stress-inducible gene for 9-cis-epoxycarotenoid dioxygenase
RT involved in abscisic acid biosynthesis under water stress in drought-
RT tolerant cowpea";
RL Plant Physiol. 123:553-562(2000).
DR EMBL: AB030293; BAB11932.1; -
SQ SEQUENCE 612 AA; 67715 MW; E3263A13293B188B CRC64;

Query Match 68.8%; Score 2168.5; DB 10; Length 612;
Best Local Similarity 68.6%; Pred. No. 3.2e-164;
Matches 423; Conservative 61; Mismatches 98; Indels 35; Gaps 8;

OY 8 AAVSGRWLGNHTTOPPLSSQSSDLSYCSLPMASRVPRKLN-VSALHTPPALHFPKO- 65
DB 3 SSASNTWFMATLPSPFKDLPTS--SPYNLPL--RKTISSSTTTCISLOT--LHFPKQY 56
OY 66 ----SSNSPAIYVKKAKKSNT-----KOMLFORAAAALDAA 100
DB 57 OPTSTSTATTTPTPIKTTITTTTPRETNPISDTNQPLQKWNFLQKAAATVLDLV 116
OY 101 BEFLVSHKHLPLKTDADPSVOIAGNFAPVNEQPYARNLPVVGKLPDSIKGYVVRNGANP 160
DB 117 ETALVSHKRNHLPLKTDADPAVOIAGNFAVPPEHAADQGLPVVGKLPKCIDGYVVRNGANP 176
OY 161 LHEPYTGHFFPDGDMVAHVAFKFEHGSASACRFQTQNRFOERQUGRPVFPKAIQELGH 220
DB 177 LYEPVAGHHFFPDGDMVAHVAFKFEHGSASACRFQTQNRFOERQUGRPVFPKAIQELGH 236
OY 221 TGIARLMFLYARAAGIYDPAHGTGVANAGLYFNGRLLAAMEDDLPRYOVTTPNGDLKT 280
DB 227 SGIAIRLLFYAAGLPGVDSQGMVANAAGLYFNNHLLLAAMEDDLPRYHVRTTPNGDLTT 296
OY 281 VGRFPDGLSESTMIAPKRVDPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 340
DB 297 VGRYPDGLSKSTMIAPKRVDPESGELFALSYDVVSKPLKFFRSPOGTSKPDVEIOLD 356
OY 341 OPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKKNVAREGILDKYAEDSSNIK 400
DB 357 EPTMMHDAFAITENFVVPDQOVVFKLPEMIRGSPVYVDKKNVAREGILDKYAEDSSNIK 416
OY 401 WIDADPCFCFHLNMAWEEETDEVVVIGSCMTPPSIFNESDENLKSYLEIRLNLKTGE 460
DB 417 WIDADPCFCFHLNMAWEEETDEVVVIGSCMTPPSIFNESDENLKSYLEIRLNLKTGE 476
OY 461 STRRPILSNEDQOVULEAGVNRNMLGRKTKPAPYALALAEPMKVSQGFPAKVDLTGGEVKKH 520
DB 477 STRRPILSNEDQOVULEAGVNRNMLGRKTKPAPYALALAEPMKVSQGFPAKVDLTGGEVKKH 535

[illegible]

RESULT	7	
Q9M6E8		
ID	Q9M6E8	PRELIMINARY;
Q9M6E8		PRT; 615 AA

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
GN NCEB1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota: Viridiplantae: Streptophyta: Tracheophyta:
OC Spermatophytes: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC eucosids I; Fabales: Fabaceae: Papilionoideae, Phaseoleae, Phaseolus
OX NCBI_Taxid:3685;

RP SEQUENCE FROM N.A.
RC STRAIN-CV, TOP CROP:
RX MEDLINE=20079657; Pubmed=10611388;
RA Olin X¹; Zeevaert J A;
RT "The 9-cis-epoxycarotenoid cleavage reaction is the key regulatory
step of abscisic acid biosynthesis in water-stressed bean.";
RD Proc. Natl. Acad. Sci. U.S.A. 96:13534-13561(1999).

KW Dioxigenase.
SQ SEQUENCE 615 AA; 68075 MW; 0CC10F862D7DE130 CRC64;

Query Match: 68.7%; Score 2165; DB 10; Length 615;

Best Local Similarity 69.28; Pred. NO. 6e-164;
Matches 426; Conservative 59; Mismatches 95; Indels 36; Gaps 9;

[illegible]

QY 65 Q-----SSNSPAIVKPKAKE-----SNTKQ-----MNLFQRAAALDAAE 101

```

Dbb 61 QYPTSTSTTTPRIPKPTTTTTTTPHRETKPPLSDTKOPKKNWLOKAAATGIDMVE 120
Ow 102 GEIVSUKET UBT PWTADBSOTACNBPVNEOCHUBN PWTCTV DSGTGVCHVNSAND 161

```

```

121 TALVSHESHKHPLEPKTADPKVQIAGNEAFVPEREHADQALPVYGVKIPKCIDGVYVRNGANPL 180

```

Dy 162 НЕРУТГННННFDGGGMMVNAVAFENGHSASTACRPTQTNRFVDEROLGRVPFKATIGELHGHNT 221
:::|||||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 181 YRPVAGNHFFDGGGMVNVAHVETNGASSTACRPFEMFQRIADKRSIGRVDFDKATCTTGHS 240

QY 222 G1ARLMLEFARAAGIYDPAHGTVANAGLVYENGRLAMSEDDLPYOVQITPBGDLKTV 281

DB 241 G1ARLLLFYARSLFQLVDSHGMVANAGLVYFNNHLLAMSEDDLPYHVRITSNGDLTV 300

```
Db      301 GRYDENGQLNSTMAHPKIDPVNGDLHALSYDVQKPYLKFRESADGVKSPVEIPLKE   360
```

QY 342 PPMNHDFALTEINFVVPDDQWFKLPEIRIGGSPVYDKNKVAAGFILDKYAEDSSNIR 401

Db 361 PPMNHDFALTEINFVVPDDQWFKLPEIRITGGSPVYDKNKSTSEGILDKNAKDANMRV 420

402 IDAPDFCFHLMNAMEEPETDEVVIVISSCMTPPDISIFNESEDNLKSVLSIIRLNKTGES 461

DB 421 IDAPECFCFLHNAMEEPEPDEIVVIGSCMTPADSIFNECEDSLKSVSEIRLNRRTGKS 480

```

0Y 462 TRRPITISNEDDOVNLACAMVRNNLGRKTRFAYTALAEPMPKVSGFAYDLTTGEVKKH 521
Db 481 TRRPITIS -DAQOVNLACAMVRNNLGRTOFAYTALAEPMKVSGFAYDLFSGEYVKM 539
0Y 522 YGDNNYGEPEPLFPGEGGEEDGYTLCPVHDEKTKWSELOVNVANSI.EVEATYKLPSPRP 581
Db 540 YGEEKFGEPEPLFPL-NGEEBGDGYTLAEVHDEKEMKSELOIVNANQNLKLEASITKLPSPRP 598
0Y 582 YGFHGTFFIGADDLAKO 597
Db 599 YGFHGTFFTHSKDLRKO 614

```

RESULT	8
Q9AXZ4	
ID	Q9AXZ4
AC	Q9AXZ4
PRELIMINARY;	
PRT;	625 AA

DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
 DE 9-CIS-EPOXYCAROTENOID DIOXYGENASE.
 GN NCED3.
 OS Persea americana (Avocado).
 OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 OC Spermatophyta, Magnoliophyta, Laurales, Lauraceae, Persea.
 OX NCBI_TaxID=3435;

RP SEQUENCE FROM N.A.
RC STRAIN-CV, LULA;
RA Chernys J., zeevaart J.A.D.;
RT "Abscisic acid in avocado fruit";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases

DR InterPro: IPR000221; protamine_P1.
DR PROSITE: PS00048; PROTAMINE_P1; UNKNOWN_1.
KW Dioxigenase.
SQ SEQUENCE 625 AA; 69724 MW; A56F33042D5F2CB6 CRC64;

Query Match	65.6%;	Score 2066.5;	DB 10;	Length 625;
-------------	--------	---------------	--------	-------------

Best Local Similarity: 64.9%; Pred. No. 4.4e-156;
Matches 392; Conservative 74; Mismatches 93; Indels 45; Gaps 6;

QY 38 LPMASRYTRKLNVSALHTPP-----ALHFP-----KQSS 67
||: :|| | | |
Pb 39 TRICKERCPDKRFTFVKNMNETTGGCGGCHDCCCAAAUUIIPKQWZCGRNHCSTZGKCEET 68

68 NSPAIVVPRKAKESNTKO-----MNLFORAAALDAAEGFLVSH--EKLHPLPKT 116
OY

Db 82 LPTSIEKNPGSHQVKTDDSGPNRVGPNNNI FORTAAEALDAIEKLIARVLERRHPLPKT 141

Or 117 ADREVOTAGNEPVDNVEQVBANI PVIYKRT DSKTCGVNBGANPI HEDIVTCHEEEEDGGM 176

```

142 ADEVOIAGNEAPVAENHPYHGIRPVAGRIPRCLDGVYVRNGANPLERPIAGNHFFDGDGM 201

```

Dh 202 : 177 VAAVKEHGSAACYACFTQGNRFVQDERGRPFVKRAIGELGHGTGIARLMLEFYARAAG 236

OY 237 IVDPAHGTGVANAGLVYFENGRLAMSEDDLRYQVQTTPNGDKTYGRFDPDGLESTMA 296
:
:

Db 262 LVNADEGVANAGLVYFNRRLLAMSEDDLPYHVRITTPSGDLKTVGRHDEFNQLRSSMIA 321

OY 297 HPRVDPESGELFALSYDVNSKPYLKYFRSPSDGTSRSPDVEIOLDOPITMWHDEAITEENFVV 356

Db 322 HPKIDPEGGELESLSYDVARKRPYLKFFHAPDGMKSPDVEILPDRPTMIDFAITENFV 381

Db 382 IPDQOVFKLEEMIRGSSPVYDKNKTSTRFGLPKKAPDASEMIWVADAPDCFEHLWNW 441

```
QY      417 EEPETDEVVVIGSCMTPPDSIFNESDENLKVLSLETRLNLTGESTRRPIISNEDQVNL 476
```

Db 442 EEPSEGVVVGSCMTPDPSIFENENEESLKSILTEIRLINTRTGESTRRTII-DPOKPLNL 500
Oy 477 EAGVNRNMLGKRTKFAYLALAEPPPKVSGFAKVDLTGTGEVKKHLXGDRYGEPLTLP 536
Db 501 EAGVNRNMLGKRTKFAYLALAEPPPKVSGIAKVDLTGTGEVNRFPYGGEGFGEPTFIR 560
Oy 537 E--GGEDEGYILCFVHDEKTKSELQIVNAVSELEAVTKLPSPVPGFHTGIGAD 593
Db 561 EPSTGSEDDGYVVSFMHDEKTSRSEILILNMMNRMLEASVMLPSRVPGFHTGIFISSRD 620
Oy 594 IAKO 597
Db 621 IAKO 624

RESULT 9

Oy 09C6Z1 PRELIMINARY: PRT: 589 AA.
AC 09C6Z1:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 9-CIS-EPOXYCAROTENOID DIOXYGENASE, PUTATIVE.
GN T2H7.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RM SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altarfi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Milttscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Uterbeck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana."
RL Nature 408:816-820(2000).
DR EMBL: AC074176; AAG50855.1; -
KM Dioxigenase.
SQ SEQUENCE 589 AA: 65336 MW: 6A79B2BF7EA8179 CRC64;

Query Match 64.9%; Score 2044.5; DB 10; Length 589;
Best Local Similarity 68.1%; Pred. No. 2.3e-154;
Matches 390; Conservative 69; Mismatches 101; Indels 13; Gaps 4;

Oy 27 SSSDLSYCSLPMASRYTRKLNVSALHTPPALHPKQSSPAIVVPKAKESNTKQM 86
Db 27 SPSSSVSFNTKPR---RRKLSANSVSDTPNLNPNPNSPNLIT-----PEKTSWM 76
Oy 87 NLFORAAALDAAGFLVSHKHLPLKTPADPSVOIAGNFAVNEQPRRLPVVGKLP 146
Db 77 NLFORAAALDAAGFLVSHKHLPLKTPADPSVOIAGNFAVNEQPRRLPVVGKLP 136
Oy 147 DSFKGVYNGANPLHEPVTGHFFPGDGMVAHVKEHGSASACFTGTNFRVOROLG 206
Db 137 DCTIDGYLNGANPLHEPVSYGHHLFDGDMVAHVKITNDASTSCTFTETRLVDEKQLG 196

Oy 207 RPPFKRAIGELHGTGIARLMLFYARAAAGIVDPAGHTGVANAGLYVFNGLLAMSDDL 266
Db 197 SPFPKRAIGELHGSIGIARLMLFYARGLELGLHNKNGTGVANAAGLYVFNHRLAMSDDL 256
Oy 267 PYOVQITPBGDLKTYGRPRFDQLESITMLAHKVPDESELEFALSYDVYSKRYLKYFRS 326
Db 257 PYOVRTDNGDLETTIGRFDFDQLSSAMIAHKIPDVTLELALSVDVKKRYLKYFRS 316
Oy 327 PGCTSPDVEIOLDDPTMMHDAITENFVVPDQOVFPKLGSPVVDKKNVAF 386
Db 317 PEGEKSPPDVEIPLASPTMMHDAITENFVITPDQOVFLSMFLKSPVYKIDGKISNF 376
Oy 387 GILDKYAEDSSNIKIDADCEFHLMAMEPEDEVVYVIGSCMTPDPSIFNESDENLK 446
Db 377 GILPRNAKASBMVWVESPEFCFHLMMNAMESPERDEVVYVIGSCMTPADSIENECDEQLN 436
Oy 447 SVLSEIRLNLKGTGSTRRTIISNEDQVVLGAVNRNMLGKRTKFAYLALAEPPPKVSG 506
Db 437 SVLSEIRLNLKGTGSTRRTIIPG-SVQNMLEAGVNRNMLGKRTKFAYLALAEPPPKVSG 495
Oy 507 FAKVDLTGTGEVKKHLXGDRYGEPLTLP--EGGEDEGYILCFVHDEKTKSELQIVN 564
Db 496 FAKVDLTGTGEVKNHRYGKRTKGEPTFPLPGLESDGEDDGYIMSYHDESESELHYN 555
Oy 565 AVSLEVEATVKLPSPVPGFHTGIGADLAKO 597
Db 556 AVTLELEATVKLPSPVPGFHTGIFVNSADMLNQ 588

RESULT 10

ID 049505 PRELIMINARY: PRT: 583 AA.
AC 049505:
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NEOXANTHIN CLEAVAGE ENZYME-LIKE PROTEIN.
GN F28J12.10 OR AT4G18350.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RM SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K., Schueler C.,
RL Submitted (Feb-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RM SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RM SEQUENCE FROM N.A.
RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RM SEQUENCE FROM N.A.
EU Arabidopsis sequencing project;
RL Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL021710; CA16715.1; -
DR EMBL: AL161548; CAB78837.1; -
DR InterPro: IPR004294; RPE65.
pfam: PF03055; RPE65.1
SQ SEQUENCE 583 AA: 65066 MW: A138F93542E50852 CRC64;

Query Match 63.2%; Score 1991; DB 10; Length 583;
Best Local Similarity 64.5%; Pred. No. 4.1e-150;
Matches 380; Conservative 82; Mismatches 105; Indels 22; Gaps 6;

Oy 16 GGNHTQPLSSQSDLSY--CSSLPMASTRVTRKLNVSALHTPPALHPKQSSNSPAIV 73

```
Db 11 GGIKTPW-----OAOIDLFRPIKROPKVYKCYOJDVTE-LTKRKQLEFPRTTAPR--- 62
OY 74 VPKAKESNTKOMNLFORAAAAALDAAGFLVSHKHLPLPTAPSVQIACNFPAPVNO 133
Db 63 -----QHPLNLNLFQKAAALAIIDAERALSISHEODSLPTAPRVOIAIINYSVPVS 116
OY 134 PYRNLPPVVGKLPDSIKGYVYRNGANPLHEPYTGHFFEDGDGMVAHVKEHSASACRF 193
Db 117 SVRRNLVTEGTLPCDIDGVIYIRNGANPMPEPTAGHHLFDGDGMVAHVKTINSASACRF 176
OY 194 TQTNFVQERQIGRPVFPKATIGELHGTGIAFLMLFYARAAAGIYDPAHGTVANAGLVY 253
Db 177 TKTERLYOKEKRGVFPKATIGELHGTGIAFLMLFYARAGLCGLINONGVAVANAGLVY 236
OY 254 FNGRLLASSEDLPYOVOYTPNGDLKTVGRRPDGQLESTMIAPHKVDESELPALSLVD 313
Db 237 FNNRLLASSEDLPYQKLTQTDGIVGRYDGDGLKSAMTAHPKLDVTKELHSLTD 296
OY 314 VYSKPYLKYFRSPDGTSKSPDVEIQLDQPTMAHDFALTENFVVPDQVYFKLPEMIRNG 373
Db 297 VYKRPYLYKFRSPDGVSPELEIPELPTMIDFAITENFVYIPDQVYFKLGEMISK 356
OY 374 SPVYVDKKNVAFGLIDKYAEDSSNKKMIDAPDCFCFLMANMBEPDEVVYVIGSCMT 433
Db 357 SPVYVDGEKVSRLGIMPXKATEASQIIVNSPETFCEFLMANMBEPDEIIVIGSCMSP 416
OY 434 PSINSESDENKSVLSEIRLNLKTGESTRRLIISNEDQVNLFGMVNRNMLGKRTKFA 493
Db 417 ADSINERDESLRSVLSERIRNLKTRKTRSLVNE--VLELIGMVNRNMLGKRTKFA 474
OY 494 YLALAEPMKVSYGFAKVDLTGTGEVKKHLGDNRYGGEPLLEPGEG---EEDEGYILCF 549
Db 475 FLAIAYPMKVSYGFAKVDLTGTGEVKKHLGDNRYGGEPLLEPGEG---EEDEGYILCF 534
OY 550 VDEETWSELOIVANVSLVEFATYVKLPSRPVYGFHGTGADIDIAKOV 598
Db 535 VDEETWSELOIVANVSLVEFATYVKLPSRPVYGFHGTGADIDIAKOV 563

RESULT 11
OYAX23 PRELIMINARY: PRT; 569 AA.
ID 09AX23;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DE 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 9-CIS-EPOXYCANOTENOID DIOXIGENASE.
GN NCEB1.
OS Persea americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; laurales; Lauraceae; Persea.
OX NCBI_TaxID=34435;
RN [1]
RC STRAIN=CV. LULA;
RA Cherys J., Zeevaert J.A.D.;
RT "Abscisic acid biosynthesis in avocado fruit.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF224672; AAK00632.1; -.
KM Dioxigenase.
SQ SEQUENCE 569 AA; 63078 MW; CD12075EE9022461 CRC64;
```

Query Match 62.8%; Score 1977; DB 10; Length 569;
Best Local Similarity 65.9%; Pred. No. 5, 2e-145;
Matches 373; Conservative 76; Mismatches 97; Indels 20; Gaps 5;

```
OY 46 RKLNVSSALHTPPALHFPK-----OSSNPAIVYKPKAKSNTKOMNLFORAAA 94
Db 9 KFTTTHSSLSHSSPVHLPLRLTLTTTTLPLHEKSORELGILL-----QEPNPAKNNFFORAAA 64
OY 95 AALDAAGFLVSH--EKLHPLPKTADPSVOIAGNFAPVNEOPVRRNLPPVVGKLPDSIKGV 152
```

```
Db 65 VALDTEDESFISGVLERRHDLPTSDPAVOISGNFAPVDEHVPQHLLPVSGRIIPRLCDGV 124
OY 153 YVRNANPLHEVTPVHHNPFEDGDGMVAHVKEHSASACRFQOTRPFQOEROLGSPVPEK 212
Db 125 YVRNANPLLEPVAHHNPFEDGDGMVAHSLRGQTVASACRFETETRLVQERAIIGRPVPEK 184
OY 213 AIGELHGHGTGIRLMLFYARAAAGIYDPAHGTVANAGLVYFNGRLLASSEDLPYOVOI 272
Db 185 AIGELHGHGTGIRLMLFYARATATGLVDSSGSGVATAGLVYFNRNHLASSEDLPYHRYV 244
OY 273 TPNGDLKTVGRFDEPQOLESTMIAPHKVDESEGLFALSYDVVSKPYLYKFRFSPDGTS 332
Db 245 TSSGDLTEYGRFDEPQOLNSAMIAHPKLDPASGELFALSYNIKKRPLKFKFTSDGKKS 304
OY 333 PVEIQLDQPTMAHDFALTENFVVPDQVYFKLPEMIRNGSPVYVDKKNVAFGLIDKY 392
Db 305 PVEIQLDQPTMIDFAITENFVYIPDQVYFKLPEMIRNGSPVYVDKKNVAFGLIDKY 364
OY 393 AEDSSNKKMIDAPDCFCFLMANMBEPDEVVYVIGSCMTPPDSIFNESDENLKSVLSEI 452
Db 365 AADSNGLRMLIDAPDCFCFLMTAMBEPEITDQVYVIGSCMTPPDSIFNESDENLKSVLSEI 424
OY 453 RLNLKTGSTRRLIISNEDQVNLFGMVNRNMLGKRTKFAYLALAEPMKVSYGFAKVDL 512
Db 425 RLNLKTGSTRRLIISNEDQVNLFGMVNRNMLGKRTKFAYLALAEPMKVSYGFAKVDL 482
OY 513 TTGEVKKHLGDNRYGGEPLLEPGEG--GEDEGYILCFPHDKTKMSELOIVANVSLVE 571
Db 483 STGEVKKHLGDNRYGGEPLLEPGEG--GEDEGYILCFPHDKTKMSELOIVANVSLVE 542
OY 572 ATVKLPSRPVYGFHGTGADIDIAKOV 597
Db 543 ASVKLPSRPVYGFHGTGADIDIAKOV 568

RESULT 12
OYAX23 PRELIMINARY: PRT; 604 AA.
ID 024592;
AC 024592;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE VIVIPAROUS-14.
GN VP14.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OX NCBI_TaxID=4577;
RN [1]
RC STRAIN=M22;
RC MEDLINE=98004553; PubMed=9342392;
RA Tan B.C., Schwartz S.H., Zeevaert J.A., McCarty D.R.;
RT "Genetic control of abscisic acid biosynthesis in maize.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:12235-12240(1997).
DR EMBL; U95953; AAB62181.1; -.
DR InterPro; IPR004294; RPE65.
DR Pfam; PF03055; RPE65; 1.
SQ SEQUENCE 604 AA; 65587 MW; CBCDC4DCB85F61BE CRC64;
```

Query Match 61.3%; Score 1930; DB 10; Length 604;
Best Local Similarity 63.3%; Pred. No. 3, 2e-145;
Matches 375; Conservative 73; Mismatches 122; Indels 22; Gaps 8;

```
OY 22 PLSOSSSDLSYCSLPMASVTRKLNVSALHMP---PALHFPKSSNSPAIVV----- 74
Db 17 PARSRRARAS-----NSVRESPPRAVSSVPPAECLQAFHHPVADLPAPSKKRAALVPGHA 71
OY 75 -KPKAKESNTKOMNLFORAAAAALDA-AGFLVVS-HEKHLPLPKTADPSVOIAGNFAPVN 131
```


RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021687; CAB16706.1; -
 DR EMBL; AL161550; CAB78919.1; -
 DR EMBL; AY056789; AAL10480.1; -
 DR InterPro: IPR004294; RPE65.
 DR Pfam: PF03055; RPE65; 1.
 SQ SEQUENCE 595 AA; 65601 MW; 711EDC4AF5920B9 CRC64;

Query Match 30.7%; Score 968; DB 10; Length 595;
 Best Local Similarity 36.7%; Pred. No. 1,5e-68;
 Matches 210; Conservative 106; Mismatches 220; Indels 36; Gaps 13;

QY 42 SRVTRKLVSSALHTPPALHPKQSSNSPAIVYKPKAKESNTKQNLFRQAAAAALDAE 101
 DB 42 SPITPSPDNDRNRKPKTLH---NRTNHTLVSSPKLRPEMTLATALF---TIVEDVIN 94
 QY 102 GFLVSHKHLPLPTADSVQIAGNFAPVNOQPVARNLPV-GKLPDSIKGYVANGANP 160
 DB 95 TFDIP---FSRPSVDKRVHLSDFAFVLDLPTDCEIIGHLPISLNGAYIRNGNP 149
 QY 161 LHEPVYGHFPGDGMVAHVAFKFGHSASACRFOTQTNRFQEROLGRVFPFKAIGELGH 220
 DB 150 QFLPRGPHLEFDGDMHAIRKHNKATLCSRYKTYKYNENKQGAIVMPVFSGFNGV 209
 QY 221 T-GIARLMLFYARAAAGIVDPAHGTVANAGLVYNGRLAMSEDDLEFYQVQITPNGDLK 279
 DB 210 TASVARGALTAARVLTQYNPNVNGIGLANTSLAFPSNRLFALGESDLPYAVALTESGDIE 269
 QY 280 TVGRDPEFGOLESTMIAHPKVPESGELFALSYDVVSKPYLKYFFPSDGTGKSPVEI-Q 338
 DB 270 TIGRDEPGKGLAMSTAHKPTDPTIGETFAFRYGPV-PPFLTYFFDSAGKKQROVPLFS 328
 QY 339 LDQPTMHDFATTENFVVVPPDOQVVF--LPEMIRGSGPVYVDNKAARFGIIDLKVAED 395
 DB 329 MTSPEFLHDFATRKHAIFAETQGLGRMMDLVLEGGSPVOTDNGKTPRLGIVIRKYAGD 388
 QY 396 SSNIKWDAPDPCFHLNNAWEEPTDEVVYIGSCMPPTDIFNSDENLKSIVSEIRLN 455
 DB 389 ESEMWFEPVPGFNIIHAINADWEDDNGSVLLAPNMSIEHTLERMD-LVHALVEKVKID 447
 QY 456 LKTGSTPRPIISNEDQVNIENAGVNNMNGRKTFRAYLALAEWPPVSGAKVDLTG 515
 DB 448 LVTLGVRHNPISAR---NIDFAVINPAFLGRCSRYAAIGDMPKISGVYKLDVSKG 502
 QY 516 E-----YKHLHYGDNRYGGEPLFLRPEGG---EEDEGVIICFVHDEKTKWSELDIVANV 566
 DB 503 DRDQCTVARRKMGSCCYGGEPPFVARDPGNPAAEEDDGVVYVHDEVYTESKFLVMDAK 562
 QY 567 S--LEVEATVKLPSPVRYGFHGTFIGADDLAK 596
 DB 563 SPELEIYAAVRLPRRVYGFHGLFYKESDLNK 594

RESULT 15
 Q94IR2 PRELIMINARY; PRT; 543 AA.
 AC Q94IR2;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE CAROTENOID 9,10-9',10' CLEAVAGE DIOXYGENASE.
 GN CCD1.
 OS Phaseolus vulgaris (Kidney bean) (French bean).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolae; Phaseolus.
 OC NCBI_taxid=3885;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Schwartz S.H., Olin X., Zeevaert J.A.D.;
 RT "Characterization of a Novel Carotenoid Cleavage Dioxygenase from
 RT plants";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY029525; AAK38744.1; -
 KW Dioxygenase.
 SQ SEQUENCE 543 AA; 61100 MW; F22C9883A05325F7 CRC64;

Query Match 30.5%; Score 959.5; DB 10; Length 543;
 Best Local Similarity 38.2%; Pred. No. 6,3e-68;
 Matches 208; Conservative 99; Mismatches 181; Indels 57; Gaps 14;

QY 98 DAEEFLVSHKHLPL-----KTAD-----PSVQINGNAPVNEQVVR 136
 DB 8 NGAEGGLV---KVDPKPNFGSSKYIDLLKLLVLFYDSSLPHNYLTFGNFGVTEPTPT 64
 QY 137 RNLPVYGLKPDISIKGVYRNGANPLHEPVYGHFPGDGMVAHVAFKFGHSASACRFOTQ 196
 DB 65 KDLPRYKHLPRCLNGEFYRVGPNPKFARVAGTHWFDGDMHGLRIKDGKATYVSREYET 124
 QY 197 NREYEROLGRPVPRPKAIGELHGTGIARLMLEYARAAAGIVDPAHGTGVANAGLVYENG 256
 DB 125 SRLKQEEYFGRSKFMK-IGDLKGLGLMVNIMHMLTKLVLDLSYGGGTTVTALVYHHG 183
 QY 257 RLAMSEDDLPYOVQITPNGDLKTYGRFDFGQLESTMIAHHPKVPESGELFALSYDVS 316
 DB 184 KILALSEADKRPALKEVEEDGDDQLTGLMDYDKRLGHSFTAHKVPFTGEMFSFGY-AHT 242
 QY 317 KPYLYFFRSPDGTSPVPEIQLDQPTMHDFATTENFVVVPPDOQVVF--LPEMIRGSGPV 376
 DB 243 PPIYIYRVIYSKGVYHNDVPIITISDPIIMHDAIENVAVFMDPLIRPKMVAKNKTLI 302
 QY 377 V-YDKNKVAREGILDKYAEDESSNIKWDAPDPCFHLNNAWEEPTDEVVYIGSCMPPT 434
 DB 303 FSDSTKARFGVLPVRYAKDEQHIWFEIPLNCFIRHANNAMW--BEDEVVLTICRLQMPK 360
 QY 435 -DSI---FNESENKLSVLSERLNLKTGESRRRIISNEDQVNIENAGVNNMNGRKT 490
 DB 361 LDNVGTYOEKLENFSNLEYEMKFMKTEGASOKKLAS----TVDPVNVNENTGKRQ 415
 QY 491 KRAYLALAEPMKVSYGFPAKD-----LTGGEVYKHLV--GDNRYGGEPLFL-- 534
 DB 416 RYVGTITDLSIAKVGIIKFDLHAERPDHCKELEVGANVGGLYDLGPKFGSEAYIIRV 475
 QY 535 PEGGEDEEGYILCFVHDEKTKWSELDIVANVSEVE--ATVKLPSPVRYGFHGTFIGAD 592
 DB 476 PGEISEEDDGYLVLPVHDEBNAGKSFVHYIDAKTMSADPAVAVELPNRVYGFHAFVTEE 535
 QY 593 DLAKQ 597
 DB 536 QLOEQ 540

Search completed: July 19, 2002, 10:15:43
 Job time: 235 sec

PT -
XX Claim 3; Fig 9; 101pp; English.
PS
XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCD3 protein.
CC The AtNCD3 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPD65 (Copea Responsive to Dehydration)
CC gene isolated from copea plant as a probe.
XX
SQ Sequence 599 AA;

Query Match 100.0%; Score 3150; DB 22; Length 599;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MASFRATAVSGRMVGNHTOPPLSSQSDLSYCSLPMASRVTKLVSSALHTPPAL 60
DB 1 masfrataavsgwvlgntqppissqsdlsysssipmarvtrklnvssalhtppal 60
OY 61 HPPKOSNSPAIVVPPKAKESNTKOMNLFQRAAAALDAEEFLVSHKPLHPLPTADS 120
DB 61 hfpkossnspsaivvpkakesntkqnmllfqrataaaldaeeflvshkplhplptads 120
OY 121 VOIAGNFAPVNDQPYRRMLPVVGLPDSIKGYVRNGANPLHEPVYGHFFDGDCGVNAV 180
DB 121 vqiagnfapvndqpyrrmlpvgklpdsikgyvrnganplhepvtyghffdgdcgvnav 180
OY 181 KEHSASASACFTOTNRVQEROLGRPVFKALIGELHGHHTIARMLFYARAAAGIYDP 240
DB 181 kshsasasacftotnrvoergolgrpvfkalgelghhtiarmlfyaraagiydp 240
OY 241 AHGTAVANGLVYFNGRLAMSEDDLPOVOITPNGDILKTVRPPDDGLESTMIAPKV 300
DB 241 ahgtavanaqlvyfngriilameddlpyvqiltpngdlktvgrfdtdgqlescmahpkv 300
OY 301 DPESGEFALSYDVVSKPYLKYFRSPDGTKSPDVEIQIDQPTMMHDAITENFVVPDQ 360
DB 301 dpesgefalsydvvskpylkyfrfspdgtkspdveiqidqptmmhdaitefvpvdpq 360
OY 361 QVVEFLPEMIRGSPVYVDKNKVARFGILDKYAEDSSNRTKWDADCFEHLMMNEPE 420
DB 361 qvveflpemirgspvyvdknkvarfgildkyaedssnrtkwdadcfefhlmwnepe 420
OY 421 TDEVVVISGCMPPDISIFNESEDENLKVLSERLNLKTGESTRRPPISEDDQVNLKGM 480
DB 421 tdevvvisgcmppdisifnesdenlksvlserlnlktgestrrppiseddqvnlkgm 480
OY 481 VNRNMLGRKTRFAYLALAEPMKVSFGFAKVDLTJTEVKKHLYGDNRYGGEPLFLPGE 540
DB 481 vnrmnlgrktrfaylalaepmkvsfgafkvdltjtevkkhlygdnryggeplflpge 540
OY 541 EDEGIIICFVHDEKTKWSLQIVNAVSLSEVETVKLPSPVPGHGTPTGADDLAKOV 599
DB 541 edegiiicfvhdektkwslqivnavslsevatvklpsrvpygftigaddlakqv 599

RESULT 2
AAE04789
ID AAE04789 standard; Protein: 605 AA.

XX
AC AAE04789;
XX
DT 10-SEP-2001 (first entry)
XX
DE Lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1.
XX
KW Tomato; neoxanthin cleavage enzyme; LENCED1; abscisic acid; ABA;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; herbicide.
XX
OS Lycopersicon esculentum.
XX
PN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Tuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI; 2001-400081/43.
XX
DR N-PSDB; AAD09401.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
PS
XX
PS Claim 3; Fig 2; 101pp; English.
XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Lycopersicon esculentum neoxanthin cleavage enzyme, LENCED1 protein
CC related to the invention.
XX
SQ Sequence 605 AA;

Query Match 72.4%; Score 2280.5; DB 22; Length 605;
Best Local Similarity 71.7%; Pred. No. 4.4e-222;
Matches 435; Conservative 63; Mismatches 90; Indels 19; Gaps 6;

OY 5 TATAVSGRMVGNHTOPPLSSQSDLSYCS--SLPMASRVTKLVSSALHTPPALH 61
DB 3 tttatntwtl----ktlmpsskefgyfaasistllknghrsglnlnsdlgappilh 57
OY 62 FPKOSSN--SPA--IVVPPKAKESN----TKOMLQFORAAALDAEGLFVSEKFLHP 112
DB 58 fpkossnyqltknmtlspkqgnmssstcknllvqkaamaidaesatltklelehp 117
OY 113 LPTADPSVOIAGNFAVNEOPVRRNLDPYVKLPDSIKGYVRNGANPLHEPVYGHFFD 172
DB 118 lptadpsvqiagnfavneopvrrnldpyvklpdkvcvgvyvrnganpllefpvtaghhff 177
OY 173 GGGWHAHVAFEHGSASVYACRFQTQNRVQEROLGRPVFKALIGELHGHHTIARMLFYAR 232
DB 178 gggwvhavgyfkgssasvycrfqterlvgekalgprvpfkalgelghghsjarlmlfyar 237

OY	233	AAGAGIVDPAHGCVANAGLVTFNGRLIAMSDDLPVQVOITTPRGDLKTYGRFPDQLES	292
		: : :	
Dd	238	gllglvdnshkgtcvanaagltvymnrllamseddlyphvvvtcpgdlktegrtfdqgalks	297
OY	293	TIIAHPKVDPEEGEFPFALSYDVASKRPLYLKYFRFSPGCTKSPPVEIOILDPTMHHDAITE	352
Dd	298	tmiatpkidrvpsgelifalsydvrlqkplylkyfrtskrgeksndveipvedctmhdfaitte	357
OY	353	NFVVVPDQQVKEFLDEMRIGSGSPVVYDNKKVAFFGLDKRYAEDSSNIKWIDABDCFEHL	412
Dd	358	nfvvpdqgvvfekmsemirgspsvyvdknkvsrfglyldkyakgdslkwveydcfcfhl	417
OY	413	WNAAHEPERDEVVYIGSCSTPDSIFNEDENIKSVLSERLNLKTGESRRRIISMEDQ	472
Dd	418	wnaaeaaetelivvgscstlpdsifnecdeglksvlselrnlktgsktrksliempde	477
OY	473	QYNLEAGVMNRNMLTKRTKFAYLALAEPMPKVSGFAFKVDLTTEGVKNHLTGDMRRYGSEPL	532
Dd	478	qynlaegvmnrnlkgrtkeyajalaeppmkvsgfkavnlftgevекflygdnhkyggpl	537
OY	533	FLPGE--GGEEDEGYILCFVHDEKTIWKSELQIVNAVSLVEATVKILPSRPYPGFHGFTIG	590
Dd	538	fllprpnskseedgyilafvhdekewkselqlvnamslkleatvkipsrxpygfngffin	597
OY	591	ADDLAKQ 597	
Dd	598	andlanq 604	
RESULT	3		
AAB72308			
ID	AAB72308	standard; Protein; 605 AA.	
XX	AAB72308;		
AC			
XX			
DT	16-MAY-2001	(first entry)	
XX			
DE	Neoxanthin cleavage enzyme-like protein amino acid sequence.		
XX			
KW	Defence-related signalling gene; sunflower; neoxanthin cleavage enzyme;		
KW	NCE; amino acid permease; AAP; glutamic acid rich protein; GRP;		
KW	pathogen resistance; abscisic acid metabolism.		
OS	Lycopersicon esculentum.		
PN	WO200112801-A2.		
XX			
PD	22-FEB-2001.		
XX			
PF	17-AUG-2000; 2000MO-US22961.		
PR	18-AUG-1999; 99US-0149656.		
PR	23-MAY-2000; 2000US-0206405.		
XX			
PA	(PION-) PIONEER HI-BRED INT INC.		
XX	(CURA-) CURAGEN CORP.		
Pt	Bidney DL, Crasta OR, Hu X, Lu G;		
DR	WPI; 2001-211215/21.		
XX			
PT	Novel isolated defence-related signalling gene isolated from sunflower		
PT	encoding neoxanthin cleavage enzyme, amino acid permease or glutamic		
PT	acid-rich protein useful for increasing resistance of plant to a		
PT	pathogen -		
XX			
PS	Disclosure; Fig 1; 135pb; English.		
CC	This invention relates to defence-related signalling genes isolated from		
CC	the sunflower (<i>Helianthus annuus</i>). The genes encode a neoxanthin cleavage		
CC	enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich		
CC	protein (GRP). The signalling gene is useful for increasing the		

Query Match	72.4%	Score 2280.5	DB 22	Length 605
Best Local Similarity	71.7%	Pred. No. 4.4e-222		
Matches 435	Conservative 63	Mismatches 90	Indels 19	Gaps 6
5 TATAATGCGWGLGNGHTHPPLSSSGSSDLSYCS---	SLPMAKSVTRKLNSSALHTRPALH	61		
3 ttttatatltwtl-----kkltsmpsketfgfaansislknqhnrgslnssltqaprlh	57			
62 PFKOSNN--SPA--IYVKKAKESN-----TGOMNFORAAALADPAAGFLVSHKRLP	112			
58 ftkgsnnygtcpknntlsphkqgenmsssstskwlnlvqdaamaIdavesaltkhelehp	117			
113 LPTKADPSVOLAGNAPVAPNEOPVRNRLPVYGLKLPDSIKGVYVRNGCANPLNHEPVYGNHPFD	172			
118 lptkadtprvqisgnfaprvpenpvcslrgtlytlrkpcvqgvgyrtnqnpflrftqaghhfd	177			
173 GCGMNAHAKVEHGSASYSACRFQTNRNFVQERQLGRVFRKALGELNHNIGTARLMFYAR	232			
178 gqamwhavqfknsgsasyacrfteterlyvqekalgrprvfkpkaigelhghsglarlmlyar	237			
233 AAGAYIDPRAHGTGVANAGAGLVYFNGRLAAISEDLPYOVQITPENGDKTGVGRFDQGLS	292			
238 glfglvdhnskgyvnaaglvlyfnmlhmseddllprhvakvtrpctgdkltegrtdfdgqls	297			
293 TMIAPRKYDPESEGETFALSVDVYSKPYLKYFRFSPDGTSPVDELQDLPRTMWHDFATNE	352			
298 tmlaprkldpvsgeflfalsydvrlqkpylkyflfsskngexndvrepredptmhdafate	357			
353 NFVVVFDQOVFKRLPEMIRFGSGPVYDQKKVARFGLDKYAEEDSNITKIDAPDCPHL	412			
358 nfvvfpdqgvvfkmesmrltgsrpsvvydkkvsrfglldkyakdgsdlkwverpdcfcfhl	417			
413 NNAHNEPPEIDVYVVGSCMTPPDSITFNEDEDNKVSLSTRLNKLKGESTRPRTISNMO	472			
418 nnaheaeeidelvnygsctmcpdpsltnecedeglkvsylseitrlnlklygskcltrsltenpo	477			
473 QVNLEAGVNRNMLGRTKFAFAYLALAEPPKYSGFEAKVADLTGTGVKKHLYGDNRYGGEPL	532			
478 qvnleagmvrnmlkgrtketayalalaepkprksygfekvnlffgevekflygdnkkygperl	537			
533 FLTPGE--GGEEDBGVLYLFCVHDEKTKWSLQLYVNAVSLSEVATVKLPSPVPGFNGFTIG	590			
538 fltpdpskreeddyllatfvhdekewkselqivnasmllklatvklprsvrpygfhgtfin	597			
591 ADDIARQ 597				
Db 598 andlang 604				

RESULT 4

ID AAED04787 standard; Protein; 612 AA.

XX AAED04787;

```
XX 10-SEP-2001 (first entry)
DT sgIarllllyarglglvdsgsgmgvanaglvymhnlIameSeddlpyhwltpngdlit 296
XX
DE Vigna unguiculata neoxanthin cleavage enzyme, CPRD5.
XX
KM Compea; neoxanthin cleavage enzyme; abscisic acid; ABA; herbicide;
KM stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KM plant growth protectant; Compea Responsive to Dehydration; CPRD5.
XX
OS Vigna unguiculata.
XX
PN EP116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKEN ) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI; 2001-400081/43.
XX
DR N-PSDB; AAD09394.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
XX Claim 3; Fig 2; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
XX neoxanthin cleavage enzyme is introduced. The improvement of stress
XX tolerance in plants is useful for example in plant breeding. Neoxanthin
XX cleavage enzyme genes are useful for producing transgenic plants. An arid
XX land can be improved by growing transformant weed for several years and
XX then removing the weed by specifically lowering stress tolerance in the
XX weed by inducing an inducible promoter. The present sequence is
XX Vigna unguiculata neoxanthin cleavage enzyme, CPRD5 (Compea Responsive
XX to Dehydration) protein. CPRD5 gene is isolated from compea plant.
XX
SQ Sequence 612 AA:
Query Match 68.8%; Score 2168.5; DB 22; Length 612;
Best Local Similarity 68.6%; Pred. No. 1,1e-210;
Matches 423; Conservative 61; Mismatches 96; Indels 35; Gaps 8;
QY 8 AAVSGRWLGNGNHTOPPLSSQSSDLSYCSLPMASRVTRKLN-VSSALHTPPALHFKPK- 65
DB 3 ssasntwfnatlpsppfkdlpsr-sptnllpl--rkssntltcslgt---lhfpkxy 56
QY 66 ----SSNSPAIVYKPKAKKSNF-----KQNLFORAAAAALDA 100
DB 57 gptststatacttptlptlctitttppretnplsdtnqplpqkwnflqkaataldlv 116
QY 101 EGFVLSHEKHLPLPTADPSVOIAGNFAPVNPVRNLPVVGKLPDSIKGYVVRGANGP 160
DB 117 etalvshckhnpktaadprvgaigafapvpehaadqilpvvgklpkcldgvyvrmgamp 176
QY 161 LAEPVYTGHHFFDGDGMVAHAKFEHGSASAYACRTQTNRFEQEROLGRPVFPFAIGELHCH 220
DB 177 lypvaghnhffdgdgmvhavkftngaaasyacrftetqrltsqgkslgyrpfpaigslghh 236
QY 221 TGIARLMFLYARAAGIYDPNAGHTGVANAGLYFNGRLLAISEDLLPYOVITPNGDLKT 280
```

```
DB 237 sgIarllllyarglglvdsgsgmgvanaglvymhnlIameSeddlpyhwltpngdlit 296
QY 281 VGRFEDDGOLESMTIAHPVDPESGELFALSYDVSKPYLTKYFRSPDGTSPDVEIOLD 340
DB 297 vgrfdngqlnstcmahnpkldpvdgdlhalsydvlgkpylkyfrtspdgvsdpveipik 356
QY 341 QPTMMHDAITENFVVPDQOVFVKLEPMIRGSPVYDKNVARKGILDKYAEDSSNIK 400
DB 357 eptmmhdfaltenfvvpdqgvfklteml tgsppvyvknkstrfgllhknakadanmr 416
QY 401 WIDAPDCCFCHLMNAMEEPTDEVVYTGSCMTPPDSIFNESDNLKSVLSEIRLNKTE 460
DB 417 widapdcfcfhlmwaweepteevvlgscmpadslfneceeslksvlselrlnltgk 476
QY 461 STRRPRIISNEDQVNLAEAMVNRNMLGRKTFAYLALAPMPKVSGEAFVDTLTGFKKH 520
DB 477 strrpriis-daeqvnleagmvrnkjgrktqayalalaepwkpvsqfakvdlisgevkkxy 535
QY 521 LYGDNRVGEPLFLPEGGEDEGYILCFVHDEKTKWSELQIVNVAVSLVEATVKLPSPRV 580
DB 536 mygeekfsggeplflp-ngqkeddyllatfvhdekewkselqlvnaqnllksaiksipstrv 594
QY 581 PYGFHGTFTGADDLAKQ 597
DB 595 pygfngftlshskdlrkq 611
RESULT 5
AAE04782
ID AAE04782 standard; Protein; 583 AA.
AC AAE04782;
DE 10-SEP-2001 (first entry)
XX Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED1.
XX
KM Neoxanthin cleavage enzyme; AtNCED1; abscisic acid; ABA; herbicide;
KM stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KM plant growth protectant.
XX
OS Arabidopsis thaliana.
XX
PN EP116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKEN ) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI; 2001-400081/43.
XX
DR N-PSDB; AAD09394.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX
XX Claim 3; Fig 10; 101pp; English.
XX
XX The invention relates to neoxanthin cleavage enzymes and their
XX corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
XX role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
XX Neoxanthin cleavage enzyme is used for improving stress tolerance in a
XX plant when expressed in a plant cell. The invention also relates to
XX methods for increasing or decreasing stress tolerance in a plant by
XX introducing the DNA into the plant, and a transgenic plant into which a
```


OY 431 MRPPOSIFNESEBENKSVSEIRPLNKTGSESTRPPIISNDQVNL EAGVNNRMGLCKRT 430
Db 432 ||||| ||||| :|||:||||: ||| |||| | || |||||:||||:
OY 491 KPAVALALAPMPKVSQSFAYDLTTGTGVKKHLYVDNNKGEGPELPLPGCGE-----EDEGX 545
Db 491 rlyaylavlaeapwkpkesgfakedisstgcltkfeygegrfggepcrcvpmdaahprgeddy 550
OY 546 ILCEPHDEKTMKSEIQIVNAVSLVEEAATVKLPSPRVYPGEFGHTFICADILAKQ 597
Db 551 vltfvnderagtsellivnaadrlratlvqjpsrvpfvgftgtlgtgleaq 602

RESULT 7
AAB72309
ID AAB72309 standard; Protein; 604 AA.
AC AAB72309;
XX
XX 16-MAY-2001 (first entry)
DT
XX Neoxanthin cleavage enzyme-like protein amino acid sequence.
DE
XX
XX Defence-related signalling gene: sunflower; neoxanthin cleavage enzyme;
KW NCE: amino acid permease; AAP; glutamic acid rich protein; GRP;
KM pathogen resistance; abscisic acid metabolism.
XX
XX Zea mays.
OS
XX WO200112801-A2.
PN
XX
PD 22-FEB-2001.
XX
XX 17-AUG-2000; 2000WO-US22961.
PF
XX 18-AUG-1999; 99US-0149656.
PR 23-MAY-2000; 2000US-0206405.
XX
XX (PION-) PIONEER HI-BRED INT INC.
PA (CURA-) CUPAGEN CORP.
XX
PI Bidney DL, Crasta OR, Hu X, Lu G;
XX
XX WPI: 2001-211215/21.
DR
XX Novel isolated defence-related signalling gene isolated from sunflower
PT encoding neoxanthin cleavage enzyme, amino acid permease or glutamic
PR acid-rich protein useful for increasing resistance of plant to a
PT pathogen -
XX
XX
PS Disclosure: Fig 1; 135pp: English.

XX
XX This invention relates to defence-related signalling genes isolated from
CC the sunflower (*Helianthus annuus*). The genes encode a neoxanthin cleavage
CC enzyme (NCE), an amino acid permease (AAP) and a glutamic acid rich
CC protein (GRP). The signalling gene is useful for increasing the
CC resistance of a plant to a pathogen such as fungus, virus, bacterium,
CC nematode or insect (e.g. European corn borer), preferably
CC *Sclerotinia spp.*, *Phoma spp.*, or *Phomopsis spp.*, by stably incorporating a
CC construct containing the gene into the genome of the plant. The gene is
CC useful for regulating gene expression in a plant, in response to a
CC stimulus such as infection with a pathogen, damage from a pathogen,
CC hydrogen peroxide, jasmonic acid, methyl jasmonate, salicylic acid,
CC oxalic acid or expression of a gene encoding oxalic acid oxidase. The
CC genes are also useful for stem-preferred regulation of gene expression in
CC a plant. The genes are useful in agriculture, particularly in the
CC breeding of crop plants with improved agronomic traits, for modifying
CC abscisic acid (ABA) metabolism and for modifying amino acid transport and
CC content in plants. The present sequence represents a neoxanthin cleavage
CC enzyme-like protein from Zea mays used in the characterisation of
XX sunflower NCE.

Seq	Sequence	604 AA:
Query Match	61.3%; Score 1930; DB 22; Length 604;	
Best Local Similarity	63.3%; Pred. No. 1.7e-16;	
Matches 375;	Conservative 73; Mismatches 122; Indels 22; Gaps 8;	
QY	22 PLLSSQSSDLSYCSLSLPMASNVTKKLNVSSALHPR-----PALHFPKQSSNPALVY-----	74
DB	17 parrsars-----nsvfsprravssvppaeclqgpfifkpradiprpsrkpaalavpgha	71
QY	75 -KPKAKESNTKOMNLFORAAALDA-KEGFLVS-HEKILHLEPKTADSVYIAGNFAVN	131
DB	72 aapraegggkqnlhfgcaaaaaldaeegfvaanylecrphnlpstadaqvlaqnfayvg	131
QY	132 EOPVRRLPLVVGKLPDSTIKGYVYRNGANPLHEPVTGHHFFEDGCGNHAVKKEHGSa-SYA	190
DB	133 erpvhelprvsgripfildgyatnrgarpcrfprvaghnlfdgqgmvhallrngaaesya	191
QY	191 GREFQTNRPFVOROLGRPFVPRKAIIGELHGHNGIARLMLFYRAAGYIDPAHGTGVANAG	250
DB	192 cftctcatlrgeralgrprfpaalgetlnghsjarialafyaraagylvpsaagfgyanaq	251
QY	251 LVYFNGRLAMSEDDLPYQVOTTPNGDLKTGYRGFDFDQLESTMIAPRKVPESGELFAL	310
DB	252 lvyfnggrllamseedlpyhvravadgdldetvgyrfdqglgcamiahpkldpatgelhal	311
QY	311 SYDVASKRYLTKYKFRSPDGTSPRYEIOLODQPTMMHDAITENFVYVVDQVVKRLPMI	370
DB	312 sydvskrylkylytrpqrqkxdsdvlelrqpcfmhdaifenfvyvrdhqvtrklqeml	371
QY	371 RCGSPVYVVDKKNVAFRGGLIDKKAEDSSNIKMWADAPDCPFILMNAMBERPETDEVVILSC	430
DB	372 rgspvrvldketrstfryglprkhaadasemawdyvdcfcflhnaeadeatgervvlygsc	431
QY	431 MTPPDSINESDENIKSVLSEIRLNLKTGCESTRPTIISNEDQYVNLEAGVNRNMLGRKT	490
DB	432 mtpadsiinesderlesvltelridatrgstftravlpr-psqgenlevgmvrnlllgres	490
QY	491 KRAYALALEPMPKVSQFPAKVDLTGVEVKHLYGDMRYGSGEPLFLPBGEGGE-----EDGXY	545
DB	491 ryayalavepmpkesgfikedselgcltkfeygegrfgyepcftvpmrpaahprgeddy	550
QY	546 ILCFVHDEKTMKSELQIVNAVSLLEVATVKLPSRYVYGFHGTFIGADDLAKQ	597
DB	551 vltfvdheragyselllvnaadirlleaetvqfprsvrpfghgtftltgqleag	602
RESULT 8		
AAE04786	AAE04786 standard; Protein: 577 AA.	
XX	AC	AAE04786;
XX	XX	10-SEP-2001 (first entry)
XX	DE	Arabidopsis thaliana neoxanthin cleavage enzyme, ATNCEDS.
XX	KM	Neoxanthin cleavage enzyme; ATNCEDS; abscisic acid; ABA; herbicide;
XX	KW	stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX	XX	plant growth protectant.
XX	OS	Arabidopsis thaliana.
XX	PN	EP1116794-A2.
XX	PD	18-JUL-2001.
XX	PF	11-JAN-2001; 2001EP-0300218.
XX	PR	13-JAN-2000; 2000JP-0010056.
XX	PR	11-JAN-2001; 2001JP-0003476.

Db	95	ftidp-----psrpsvdkhvltsadifapyldeiprtdceiingltlplslngayirngnp	149
Oy	161	LHEPVTGHHFEDGDGMVHAVKFEHGSASYACRFQTOTNRFVROLOGRPYFPAKIGELHGH	220
Db	150	gfllpgrp/hlfdgqdmllhaiklhngkatlcsryvktkynevkqsgapmpvrfsgfngv	209
Oy	221	T-GIARMLIFYRAAAGIVDPAHGCVANAGLVYINGRLAMSEDDLPYOVQITPNGDLK	279
Db	210	tasvargaltaarvltgqynpvnigilantslafsnsrlfalgesdlpyavrlltesgdl	269
Oy	280	TWGRDFPGQGLSEWIAIHPKVDPEGSELFALSYDVVSKPYLKYFPFSPDCTKSPDVEI-Q	338
Db	270	tlgrtltlrltlrltlrltlrltlrltlrltlrltlrltlrltlrltlrltlrltlrltlrl	328
Oy	339	LDQPMHMDFAITENFVVVDDQOVFK--LPEMIRGSGSPVYVDKNKVARFGILDKTAE	395
Db	329	mcspsflndfalkthkhaelqldmmmlvlyegspsvgtcnkcpriydpkysgd	388
Oy	336	SSNIKWIDAPDCECFHLMANWEEPTDEVVYIGSCWTPPDSIFNSDENLKSVLSEIRLN	455
Db	389	esemwfwefvgfnllhainawdedgnsavllaplmslhtlermd-lvhalvekvkld	447
Oy	456	LKTGSTRRPRIISNDDQVNLNLAGVNNRMLOGRKRFAYLALAEWPKVSGAKADLTIG	515
Db	448	lvtglvtrhplisar-----nldfavlmpaflgrcsriryaaigdpmpksgvkvldvskg	502
Oy	516	E-----VKKHLTYGDNRYGGEPLFLPGEGG---EEDEGFIICFVHDEKTKWSELDIVNAV	566
Db	503	dtddctvarrmgysgcyg99effvarcdpnpneaeddgyvvyvndevtlgesklfvmak	562
Oy	567	S--LVEEATVKLPSRNPYGVGHGTFIGADDLAK	596
Db	563	spelelvaavrllprvpygfbhglfkesdlnk	594
	RESULT	10	
	AAE04783		
ID	AAE04783	standard; Protein; 595 AA.	
XX			
AC	AAE04783;		
XX			
DT	10-SEP-2001	(first entry)	
XX			
DE	Arabidopsis thaliana	neoxanthin cleavage enzyme, AtNCE2.	
XX			
KM	Neoxanthin cleavage enzyme; AtNCE2; abscisic acid; ABA; herbicide;		
XX			
KM	stress tolerance; transgenic plant; plant breeding; antisense-therapy;		
XX			
KM	plant growth protectant.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP116794-A2.		
XX			
PD	18-JUL-2001.		
XX			
PE	11-JAN-2001; 2001EP-0300218.		
XX			
PR	13-JAN-2000; 2000JP-0010056.		
XX			
PR	11-JAN-2001; 2001JP-0003476.		
XX			
PA	(RIKE) RIKEN KK.		
XX			
PI	Iuchi S, Kobayashi M, Shinozaki K;		
XX			
DR	WPI; 2001-400081/43.		
XX			
DR	N-PSDB; AAD09395.		
XX			
PT	A DNA encoding a protein with a neoxanthin cleavage activity for		
XX	producing transgenic plants with improved or decreased stress tolerance		
XX	-		
XX	Example 10; Fig 10; 101pp; English.		

CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCD2 protein.
CC The AtNCD2 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPR65 (Cowpea Responsive to Densitration)
CC gene isolated from cowpea plant as a probe.
XX Sequence 595 AA;

Query Match	30.7%	Score 968	DB 22	Length 595
Best Local Similarity	36.7%	Pred. No. 8.3e-89		
Matches 210; Conservative	166	Mismatches 220;	Indels 36	Gaps 13

[illegible]

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144321.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144633.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145066.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146387.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160816.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 29.8%; Score 939; DB 21; Length 538;
Best Local Similarity 37.0%; Pred. No. 6-2e-86;
Matches 207; Conservative 106; Mismatches 198; Indels 48; Gaps 13;

OY 64 KOSNSPAIVPK-AKESNTKOMLFDRAAALDAAEGLVSHBKHLPLPTADPSVO 122
DB 4 klsdgsllsvhprpsksgskldllderlvklm-----hdasjplh-----y 47
OY 123 IAGNFAPV-NEQPVRRNLPVVGKLPDSIKGVVRRGANGPLHEPVGHHFFDDGDGVAHVAK 181
DB 48 lsgnflaprldecpvkdlprhgfllpeclngeflrvvgprpklfavagymhfdgdmhgvrr 107
OY 182 FEHGSASYACRPTQTNRFVQEROLGRPVPRKATIGELGHTGJARLMLFYARAAGIVDPA 241
DB 108 lkdqkatlystryvksrllkqekffgaakfmk-igdlkgffglmwlniqdlrtklklidnt 166
OY 242 HGTGVANAGLVYFNGRLAMSEDDLRYOVOTPNNGDLKTVGRFDDGOLESTMAHPKVD 301
DB 167 yngntantalyhghklalgaedkpyvklvledgdllqrlgldykrlthstahpkrvd 226
OY 302 PESGEFALSTYDVSKPRYIKRFPSPDGTSPDVEIQLDQPTMMDFATTEFVVVPPDQ 361
DB 227 pvtgemftfygs-hcpbplylvrvlskdglmhdpvbltiseipimhdfatetalyafmdpr 285
OY 362 VVEKFLPEMIRGSGSPYV-YDKNRKVARFGIIDKYAEDSSNKKWDADPCFCHLMNAEPE 420
DB 286 mhfrpkemwkekkmlystfopkkaftfyvpryakelmlrvfelpnclflfhaanwe--e 343
OY 421 TDEVVIGSCMTPD----SIFNESDENLKSVLSEIRLNLKTGSTRRPRTISNEDQOVN 475
DB 344 edevvlltcrlempdlmwsqgvkelenfgnelyemfrfmmktsasqkksasa----- 398
OY 476 LEAGVNRNMGRKTKFATALALEPMPYSGFAKXD-----LTTGVKKHLY-- 522
DB 399 vdfprlnecytgkqkryvygtilldslavkvtgllktldlhaeactgrmlevggnlgyldl 458
OY 523 GDNRYGGEPLFLPGEGEDEDEGYILCFVHDEKTKWKSQIIVANVSLVEV--ATVKLPSRV 580

Db 459 gggrygsealyprcraeeddgyllffvndentgkscvclvldaktnsaepvavveiphrv 518
QY 581 pyGFHCTFGADDLAKQV 599
Db 519 pyGfhalvtveeqIgeqtI 537

RESULT 14

AAAG13115
ID AAAG13115 standard; Protein; 539 AA.

XX
AC AAAG13115;

XX
DT 17-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 12488.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132563.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156459.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

	Query Match	29.8%	Score 939	DB 21:	Length 539:
	Best Local Similarity	37.0%	Pred. No. 6.2e-85:		
	Matches 207:	Conservative 106:	Mismatches 199:	Indels 48:	Gaps 13:
Oy	64	KOSSSPATVYPK-AKESNTRKOMLFFORAAALADAEGLFVSHEXKHLPDKTPADPSVQ	122		
Dd	5	kldsgslfsvhprrskgfskklidlalerlvtklm-----hdasplph-----y	48		

QY 123 IAGNAPRV-NNEQVRNRNLPVYCKLRDSDIKGVYVNRGANCANLHERPVYGNHREDDGDWVNAHX 181
Db 49 IsgnIapiridreppvKdIprvhfIReclngeIvtrvgprnKfIaavayIhIdgdgmIhgyr 108
QY 182 FEHGASAYACRTQTNRNFEQEROLDRPVPYPAKIGELHGHGTGIALMLFYARAAGIADPA 241
Db 109 IkdgratIvrsryvKtsrIkqekfIgaakfImK-igIdkgfIglmnhIqIrlckIkIdnt 167
QY 242 HGTGVANAGLVYFNGHILAMSEDDLRYOYQITPNNGDLKTVGRFDFDGOLESTMIAPHKVYD 301
Db 168 YngftancatlvyhKglIalIgaadKpryIkvIedgdIglIglIdydkrltIhstfIahprkvD 227
QY 302 PEGSEFALSTPVYVASKPYKTYRFESPDGKKSPDVEITOLDOPMTMHDFATITENFVVVDDQ 361
Db 228 pvtgemfItIcIgs-hcpYpItyrIvIsKdgmhbpvritIseplImmhndfalteyalfmdIpr 286
QY 362 VYFKLEPMIRIGSGSPVY-XDKNKVAREGILDKYAEDSSNIKWIDAPDFCFHLMNAWEEPE 420
Db 287 mIfrfepemKmekkIySfdpIctkarfIvIpryakdeImIrvfelpocfIffhnanawe--e 344
QY 421 TMEVVVYSCMTPRP-----SIFNSDENLKSVLSEIRINLTKGSTRPRIISNEDQOVN 475
Db 345 edevallIrtIenpIdImwsgkvKexIenIenIemIrfmKtIgsaSqKkIsasa----- 399
QY 476 LEAGWNRNMLGRKTKFAVIALAEPPWKSGFAKVD-----LTTCGEVKKHLY-- 522
Db 400 vIfrIrineyIcKkqaryygtIIdstIakvtIglIkIdIlnhaetqKrmIevgnIglIdI 459
QY 523 GNNRYGEPFLPLPREGDEDEGYILCFVHDEKTKWSELOIVANVSLEVE--ATVKLPSRY 580
Db 460 gsgrygseaIyIpreIaeedgyIIfIvfhndentIgtKscvtIIdakItnsaepvavveIhprv 519
QY 581 PYGFGTGFIGADDLAKOVY 599
Db 520 pygfhallIvteeqIgeqtI 538

RESULT 15
AAE04785
ID AAE04785 standard; Protein; 538 AA.
XX
AC AAE04785;
XX
DT 10-SEP-2001 (first entry)
XX
DE Arabidopsis thaliana neoxanthin cleavage enzyme, AtNCED4.
XX
KW Neoxanthin cleavage enzyme; AtNCED4; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
XX plant growth protectant.
XX
OS Arabidopsis thaliana.
XX
PN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-0300218.
XX
PR 13-JAN-2000; 2000JP-0010056.
XX
PR 11-JAN-2001; 2001JP-0003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Iuchi S, Kobayashi M, Shinozaki K;
XX
DR WPI: 2001-400081/43.
XX
DR N-PSDB: AAD09397.
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance
XX

PS Example 10; Fig 10; 101bp; English.

XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key
CC role in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present sequence is
CC Arabidopsis thaliana neoxanthin cleavage enzyme. AtNCED4 protein.
CC The AtNCED4 cDNA is obtained from an Arabidopsis plant-derived cDNA
CC library using a cDNA of the CPD65 (Cowpea Responsive to Dehydration)
CC gene isolated from cowpea plant as a probe.

XX
SQ Sequence 538 AA;

Query Match 29.8%; Score 938; DB 22; Length 538;

Best Local Similarity 37.0%; Pred. No. 7.8e-86;

Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

OY 64 KQSSNSPAIVVPRK-AKESNTKOMLFORAAALDAAGFLVSHKILPRKTADPSVQ 122
DB 4 KisdgsslsivnhrpskgsfkldlrlrvkIm-----hdasjplh-----y 47
OY 123 IAGNFAPV-NEQPVNRNLNVGKLRPSIKGVYVRNGANPLHEPVTHNPFDDGWNHAVK 181
DB 48 Isgnfapirdeprkdlrvhgflpeclngefvrvgpnrpkfdavagyhwhfdgdmhgyr 107
OY 182 FEHSASVACRTQTRRFQERQLGRVPRPKAIGELHGTGIALRLMFLYARAAGIVDPA 241
DB 108 IkdgkatyvsryvktsrlkqeeffgaakImk-IgdIkqffglmrvvgqlrIkIkIdnt 166
OY 242 HGTGVANAGLVENGRFLAMSEDDLPYOQITPNGDLTKVGRFPDGOESTMIAPKVD 301
DB 167 yngntaantalyhngklalgaedkpyikvleogdlqclgldykrlthsfanpkvd 226
OY 302 PESGELFALSYDVNSKPYLKYFRFSPDGTSPDVEIQDQPTMMDFAITENFVVVPDQ 361
DB 227 pvtgemfftye-heppyltyrviskdglmhdprvptlseplmmdfatletyalfmdlp 285
OY 362 VVFRLPEMIRGSGSPVY-IDKNKVARPGILDKYAEDSSNITKWDAPDCFCFHLNMEPE 420
DB 286 mhfrpkemvkekmlkysfdpkkarfgypryakdelmrlwfeIpnclffhnanawe--e 343
OY 421 TDEVVVVIGSCMTPPD-----SIFNSDENLKSIVLSIRLNLTKGSTRRPITSNEDQVN 475
DB 344 edevvlltcrlepnoldmvsqkvkelenfgenfymrfmktysasqkIsasa----- 398
OY 476 LEAGVNRNMLGRKTKFAYALAEPPKVSQFPAKYD-----LTTGEVKKHLY-- 522
DB 399 vdfprlneqytkkqryvgvtlIdsiakvtglfkfdlhaeetgkrmlevgnIkgyld 458
OY 523 GDNRTGGEPLFLPREGGEDEGYIILCFVHDEKTKWSELDIYNAVSLVEY--ATVKLPSRY 580
DB 459 gegryysaealyvpretaeedgylIfivndentgksvcvldakImaepvaaveIphrv 518
OY 581 PYGFHGTFFGADDLAKOVY 599
DB 519 pygfthallvteeqIdgecl 537

Search completed: July 19, 2002, 10:11:42
Job time: 45 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:11:03 ; Search time 21.04 Seconds
(without alignments)
2735.623 Million cell updates/sec

Title: US-09-758-269-6
Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGNNHT.....VPYGFHGFICADLAKQVY 599

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2289	72.7	604	2 T51936	probable 9-cis-epo
2	2280.5	72.4	605	2 T07123	nine-cis-epoxycaro
3	2170.5	68.9	657	2 E96812	protein F3F9.10 [1
4	2044.5	64.9	589	2 A86425	probable 9-cis-epo
5	1991	63.2	583	2 T04531	nine-cis-epoxycaro
6	1930	61.3	604	2 T04351	viviparous-14 prot
7	968	30.7	595	2 T04438	hypothetical prote
8	938	29.8	538	2 T49193	neoxanthin cleavag
9	937	29.7	538	2 T51734	neoxanthin cleavag
10	745.5	23.7	475	2 AG1944	hypothetical prote
11	745	23.7	446	2 T17019	hypothetical prote
12	528.5	16.8	483	2 E87345	conserved hypotet
13	466	14.8	501	2 A70534	hypothetical prote
14	454.5	14.4	502	2 A70582	hypothetical prote
15	394	12.5	480	2 S76169	hypothetical prote
16	355.5	11.3	497	2 AE2341	hypothetical prote
17	352	11.2	484	2 JN0595	lignostilbene alph
18	317	10.1	480	2 S76206	lignostilbene alph
19	303.5	9.6	489	2 JCA324	lignostilbene alph
20	253	8.0	616	2 T10688	hypothetical prote
21	244	7.7	618	2 C84885	hypothetical prote
22	242	7.7	472	2 AG2417	hypothetical prote
23	238.5	7.6	556	2 F88115	protein F53C3.12 [
24	231	7.3	483	2 D87290	conserved hypotet
25	215	6.8	533	2 A47143	retinal pigment m
26	126	4.0	790	2 T34293	hypothetical prote
27	116.5	3.7	1832	2 T31113	mucin-like glycopr
28	114	3.6	870	2 S74291	hypothetical prote
29	113	3.6	466	1 IOEBV	replication initia

30	110.5	3.5	658	2 S60170	protein kinase Pak
31	108	3.4	437	2 G70019	conserved hypotet
32	107.5	3.4	658	2 T39500	serine/threonine-s
33	107	3.4	1461	2 T13157	mitotic checkpoint
34	107	3.4	26926	1 T38344	titin, cardiac mus
35	106.5	3.4	1299	2 T47182	hypothetical prote
36	106	3.4	1679	2 T15968	hypothetical prote
37	106	3.4	3562	2 A47171	chondroitin sulfat
38	105.5	3.3	2591	2 T30288	pristinamycin I sy
39	105	3.3	2504	1 A57788	fatty-acid synthas
40	104.5	3.3	623	2 T52293	MYC-related DNA bi
41	104.5	3.3	1080	2 A35088	phycobillosome link
42	103.5	3.3	394	2 AG3616	hypothetical prote
43	103	3.3	775	2 B70449	phenylalanine--trN
44	103	3.3	4845	2 T31067	BTR repeat contain
45	102.5	3.3	531	2 AH3378	cell wall degradat

ALIGNMENTS

RESULT 1
T51936
probable 9-cis-epoxycarotenoid dioxygenase [imported] - potato
C:Species: Solanum tuberosum (potato)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T51936
R:Birdidge, A.; Taylor, I.B.; Thompson, A.
Submitted to the EMBL Data Library, March 2000
A:Description: Potato putative 9-cis-epoxycarotenoid dioxygenase 1 cDNA.
A:Reference number: 225874
A:Accession: T51936
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: mRNA
A:Residues: 1-604 <B0R>
A:Cross-references: EMBL:AJ276244; PIDN:CAB76920.1
C:Genetics:
A:Gene: nced1

Query Match	72.7%	Score 2289	DB 2	Length 604
Best local similarity	71.3%	Pred. No. 1.2e-165		
Matches 432; Conservative 66; Mismatches 90; Indels 18; Gaps 5;				
QY 5 TATAVSGRWLGNNHTPPSSQSDLSYCS---SLPMASRVTRKLVSSALHTRPPLH 61				
DB 3 TTSTMTNTWT----KKPLSPSPSKKEFGFASNSISLKNQHNQSLNINSLSLAPPLH 57				
QY 62 FPKOSSN-----SPATVVKPKAKESNT---KOMLFOFAAALDAAGFLVSHKCLHPL 113				
DB 58 FPKOSSNQTQRTKSTISHPKQENNNSSISKMLVQFAAALDAVGCALTKEHLEHPL 117				
QY 114 PKTADPSVQIAGNFAPVNEQVRRNLPVYGLKLPDSIKGVYRNGANPLHEPVYTGHHFEDG 173				
DB 118 PKTADPVRVQISGNFAPVNEPVCOSLPVYTKIPKCVQGVYRNGANPLFEPTAGRHFPDG 177				
QY 174 DGMVAHVAFFEHGSASYACRFQTNRVQEBROLGRVPRFKALGELHGHGTARMLFTYARA 233				
DB 178 DGMVAHVAFFEHGSASYACRFETERFVEKALGRVPRFKALGELHGHGTARMLFTYARG 237				
QY 234 AAGIVDPAGHGTGVNAGAGLVYFNGLAMSEDDLPVOVITPNSGDKTGTGRFPDQGLEST 293				
DB 238 LFGILDRHGRGTGVNAGAGLVYFNNNRLAMSEDDLPYHVAVTPTGDKLTGRDPDQGLKST 297				
QY 294 MIAPHKVDPESGELFALSVDVYVSKPYLYKFRSPDGKSPVEIQLDQPTMMHDPATEN 353				
DB 298 MIAPHKLDVPSGELFALSVDYDIQKPYLYKFRFSKNGEKSNDEIVEEDPTMMHDPATIEK 357				
QY 354 FVVVPDQGVVFKLEPMIRIGGSPVYVDKKKVARFGLDKYADSSNIKKIDAPDQCFELW 413				
DB 358 FVILPDQGVVFKMSBMRIGGSPVYVDKKKVSFGLDKYADGSDLKWEVDQCFELW 417				
QY 414 NAMEEPETDEVVVGSCMTPPDSIFNSEDNLKSVLSERINLKTGESTRRPTISNEDQ 473				

Db	418	NAMEPELDELIVYGSCMTPPDSIFNEBCDEGLKSVLSEIRLNKTIKSTKRAIIENPBQ	477
Qy	474	VNLEAGVNNRMILGRKTRFAYLALAEWPVKVSFEARVDITTTGEVKKHLYGDNRYGEEPLF	533
Db	478	VNLEAGVNNRKLGRKTRQYAYLALAEWPVKVSFEARVDITFTGEVEKEFIYGDNKRYGEEPLF	537
Qy	534	LRGE--GGEDEGEYILCFVADERTWKSLEQIVAAVSLAEVATPKLPSRPVYGFHGTFTG	591
Db	538	LPRPNRSKEDDQITLLAFVADDEKEMISEQIVAAVMTLLKLEPAVTKLPSRPVYGFHGTFTG	597
Qy	592	DDLAKQ	597
Db	598	NDLANQ	603

RESULT 2
T07123
nine-cis-epoxycarotenoid dioxygenase - tomato
N:Alternate names: probable neoxanthin cleavage enzyme
C:Species: Lycopersicon esculentum (tomato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07123
R:Burridge, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z15934
A:Accession: T07123
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-605 <BUR>
A:Cross-references: EMBL:Z97215; PIDN:CAB10168.1

[illegible]

```

OY      533  FLPGC--GGEDDECYIICFVHDEKXTWKSELOIVANASLEWATYKLESRPYCPHGHFIC 550
Db      538  FLPRPNSEKDEEDGIIILAFVHDEKEMKSELOIVANASLKLDEATYKLESRPYCPHGHFTIN 597
OY      591  ADDILAKQ 597
Db      598  ANDILANQ 604

```

RESULT 3
 E96812
 Protein F3P9.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: E96812
 R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alcorn
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Hutzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141, MUID:21016719
 A:Accession: E96812
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-657 <STO>
 A:Cross-references: GB:AE005173; NID:98052533; PIDN:AAE71797.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F3P9.10
 A:Map position: 1

Query Match	68.99	Score 2170.5	DB 2	Length 657
Best Local Similarity	67.78	Pred. No. 1:3e-156		
Matches 419	Conservative 80	Mismatches 87	Indels 33	Gaps
QY	1	MASFATAAGVGRML-----GNNHTPOPLSSQSSDLSCSSLP-----MASRVT	45	
DB	51	MASTLLPESTSTQGLDRFTFSSSSSRKLOS-----LSFSSTLRNKKLVLPVCYSSVN	105	
QY	46	EKLNVSSALHT---PRLAHPKQSSNSPAIVYKPKAESNTKOMLFOBAAAAALDAE	101	
DB	106	KKSSVSSSLQSPTEKPSPM---KTKLDNVNLI-PKTTNQNPK-LNPQSTAAVMDAVE	160	
QY	102	GLVSHF-KLHPLEKTDAPSOVQVQGNAPVQVRRNLVYGLKPLDSIKGVYVRGAMP	160	
DB	161	NAMLSHERRRHPHKTADPAVOVIGNFFPVEKPVVNLPTVGTGPRICGVYVRGAMP	220	
QY	161	LHEPVTGHFFKDGGMVAHYVEEIGSASVACRFQVOTRPFVQEOELGPRVCPKATIGLGH	220	
DB	221	LHKPVSGHLLFDGDMVAHYFDDGVSYSACRFETENKLVQERCGRPVPEKALGELGH	280	
QY	221	TGIRALMFYARAAGIYDPAGHAGVANAAGLVYENGRLILMSEDDLYQVOTIPNGDLKT	280	
DB	281	LGIAKLMLFENRGLFGLVDPFGGIGVANAAGLVYENGHLLMSEDDLYHYKAVYQTGLET	340	
QY	281	VGRFEDQLESTMIAPHKVDPESEGELFALSYDVYSKPYLKYFRFSPDGTSPPVEIQLD	340	
DB	341	SGRVYFDQGLKSTMIAPHKIDPEIRRELFALSYDVYSKPYLKYFRFSDGKSPVEIPLD	400	
QY	341	OPTMHADPAITENFVYVDDQVYKRLPEMIRGSSPVYIDKNKARFGLDKYEDSSNIK	400	
DB	401	OPTMHADPAITENFVYVDDQVYKRLPEMIRGSSPVYIDKSKRSFGLNKNKAKDASSIQ	460	
QY	401	WTDADCCCFHLMANMEPEITDEVYVIGSCMTPDDSTFNEDEMLKVLSEIRNLKTGE	460	
DB	461	WLEPDCCFHLMANMEPEITDEVYVIGSCMTPDDSTFNEDETLQSLSEIRNLKTGE	520	
QY	461	STRRIIENEDQVVLLEGVNRNMLGKTKFAVIALAEPKPKSGAKVDLTGGEVKKH	520	

Db	521	STRRRVIS --- EQVNLLEGMVNRNLLGRKTRAYALALTEPMVKVSGFAKYDLSIEIRKY	577
Oy	521	LYGDNRRYGGEPFLFPGEGEEDGYIILCFVHDEKTKWSELOIVNAVSVLEAVTKLPSPRV	580
Db	578	IYGEKRYGGEPLFLPSGGEGEDGYIMVFVHDEKRVKSELOINAVNMKLEAVTILPSRV	637
Oy	581	PYGFHGTIFGADDLAKQYV	599
Db	638	PYGFHGTIFSKEDLSKAL	656

RESULT 4

AB6425

Probable 9-cis-epoxycarotenoid dioxygenase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: AB6425

R:Orthologs: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malli, R.; Marzalli
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719

A:Accession: AB6425

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1*589 <SMD>

A:Cross-references: GB:AE005172; NID:q11094779; PIDN:AA629711.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match	Similarity	64.9%	Score	2044.5	DB	2	Length	569
Best Local	Similarity	68.1%	Fred	No.41e-147				
Matches	390	Conservative	69	Mismatches	101	Indels	13	Gaps
4								
QY	27	SQSSDLSYSCSSLPMASSRATRKLNVSALHTPPALHPEPKROSSNSPAIVKPKAKESNTKOM	86					
DB	27	SPSSSVSTNTNKR-----RRKLSANSVSDTPRLNLFNPNPSPNPIL-----PEMDISRM	76					
QY	87	NLFQRAAAALDAAAGELVSHSEKHLHPLEKTDPSVQIAGNFAVNEQYRRNLPVYGLP	146					
DB	77	NPLQRAAALDFAETALLRERASKPLRKYDPRQISQNVAPVEOSKSLSDVGKIP	136					
QY	147	DSIKRVYRNGANPLPHEPVTHGHPFDGDMVAHVKEHGSASACRFOTNTNFFVOEROLG	206					
DB	137	DCIDGVYLRNGANPLPEFVPSGHHLFDGDMVAHAKITNDGASYSGRFETELVQEKQLG	196					
QY	207	RPVFKAIGELHGTGIGIARLMIFYARAAGIVDPAHGCGVANAGLYTNGRLLAMSEDL	266					
DB	197	SPIFKAIGELHGHGIGIARLMIFYARGLELHNRNGGVANAAGLYFHRDLAMSEDDL	256					
QY	267	PIYQVITNGDILKYVGRDFDQGLSESTMIAHPKVPPESEGLFALSVDVASKRYLYTFRPS	326					
DB	257	PIYQVAVTNGDLELTIGRRDFDQGLSSAMIAHPKIDPVYKELFALSVDVAKRYLYTFRFS	316					
QY	327	PDGTSFDPVEIOLDQPTMHMDFATITENFVVVDDQOVFKLPEMIRNGSEPVYDKKKVAF	386					
DB	317	PEGKSPDPVEIPLASPTMHMDFATITENFVVIDQOVFKLSDMELGKSPVYDKKISNRF	376					
QY	387	GLIDYVAEDSSNIKIDAAPCCFHLMANAMEPEDEVVIVISCATPPSITNEDENIK	446					
DB	377	GLIPRNADADASLMYVSESEETCFHLMANAMESPEIDEVVIVISCATPPADSTFNEDCDQLN	436					
QY	447	SVLSEIRLNLKGTGSTRRPITISNEDQVNLKAGVNRNMLGRTKFAVYLAIALEMPKVS	506					
DB	437	SVLSIRLNLKGTGSTRRTIIPG-SVQNNLEAGVNRNMLGRTKFAVYLAIALEMPKVS	495					

OY	FAKVDLTSTEVNHHFYGGKKTYGCEPFELPRGLSEJSGEDDGTIMSEFVHDESMESLHTIYN	555
Oy	507 FAKVDLTSTGVXKHLYIGDNRYRGGEPLFLPG--EGGEPEBGHLLTCVHNHEKTMKSLOLIYN	564
Db	496 FAKVDLTSTEVNHHFYGGKKTYGCEPFELPRGLSEJSGEDDGTIMSEFVHDESMESLHTIYN	555
OY	565 AVSLEVEATVKLPISRPYPVGHGTFIGADDLAKQ	597
DB	556 AVTLEAEATVKLPISRPYPVGHGTFIVNSADMILNO	588

RESULT 5
 T04531
 nine-cis-epoxycarotenoid dioxygenase homolog F28J12.10 - Arabidopsis thaliana
 M:Alternate names: hypothetical protein T9A21.200
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
 C:Accession: T04531: T04937
 R:Byan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
 submitted to the Protein Sequence Database, February 1998
 A:Reference number: Z15377
 A:Accession: T04531
 A:Molecule type: DNA
 A:Residues: 1-583 <BEV>
 A:Cross-references: EMBL:AL021710
 R:Byan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.
 submitted to the Protein Sequence Database, February 1999
 A:Reference number: Z15390
 A:Accession: T04937
 A:Molecule type: DNA
 A:Residues: 1-377 <BEV>
 A:Cross-references: EMBL:AL021713
 A:Experimental source: cultivar Columbia; BAC clone T9A21
 C:Genetics:
 A:Map position: 4
 A:Note: F28J12.10; T9A21.200

[illegible]

[illegible][illegible]

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49193
R:Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke, K.
submitted to the Protein Sequence Database, April 2000
A:Reference number: 225018
A:Accession: T49193
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1538 <R1E>
A:Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.150
C:Genetics:
A:Gene: ATSP:MAA21.150
A:Map position: 3
A:Introns: 72/3; 96/2; 101/2; 138/3; 172/1; 193/1; 230/1; 291/3; 338/1; 377/2; 411/2; 49

Query Match 29.8%; Score 938; DB 2; Length 538;
Best Local Similarity 37.0%; Pred. No. 3.8e-63;
Matches 207; Conservative 105; Mismatches 199; Indels 48; Gaps 13;

QY 64 KQSSNSPAIYVKPK-AKESNTKOMNLFORAAALDAEAGFLVSHKRLPLPKTAPSPVQ 122
DB 4 KLSDGSSIIIVHRRPSKGFSSKLIDLLERLVVKLM-----HDASLPILH-----Y 47
QY 123 IAGNFAPV-NEOPVRRNLPVVVKLPDSIKGYVRNGANPLHEPVYGNHFFDGDGMVAHV 181
DB 48 LSGNFAPIRDETPPVKDLPHVGFPLECNGEYVRVGNPKFPAVAGYHMFDDGMHGV 107
QY 182 FEHGSASACRFQTNFVQERQLGRPVFPKALIGLHGHTGIARLMLFYARAAAGIVDPA 241
DB 108 IKDGKATVYSRVYKTSKLKEEFGAAMFK-IGDLKGFFGLMVAWVQDLRTKLKILDNT 166
QY 242 HGTGVNAGIYVYVNGRLLAMSEDDLPRYOQITPNGLDKTVGSRFDFDGOLESTMIAPKVD 301
DB 167 YNGNTATVYVHNGKLLALQEADKPRYIKVLEDDLTGLGIIYDKRLTHSTFAPKVD 226
QY 302 PESGELFALSYDVVSKPYLKFRFSPDGTSPDVEIOLDPIMHDFALTENFVVPDQ 361
DB 227 PYVGEMETFEYS-HTPPYLFYRVISKDGIIMDPVPTISEPIMHDFALTETVAIFMDLP 285
QY 362 VYFKLPEMTIGGSPV-YDKNKVAFEGILDKYAEDSSNIKWIDAPDFCFHLMNAEPE 420
DB 286 MHRPRPEMYEKMKIYSEDFPKKARFGVLPYAKDELMIRWELPNCFTFHNAME--E 343
QY 421 TDEYVYIGSMTPPD-----SIFNESDENLKSVLSEIRNLKTGSTRPIISNEQOVN 475
DB 344 EDEYVVLITCLRNPDLDLMVSGVKREKLENGNEIYEMRNMTGSASOKLSA----- 368
QY 476 LEAGVYVNRNMLGKRTKFAVYALALAPWPKVSGFAKVD-----LTGGEKKHLY-- 522
DB 399 VDEPRINCEYTGKKQRRVYVGTILDSIAKVTGIKIKDLNAEATGRMLVGGVIGKGYLD 458
QY 523 GDNRYGGEPLFLPEGGEDEBGYLLCFVHDEKTYMKSQIYNAVSLEVE--ATVKLPSPRV 580
DB 459 GGEKRYGSEALIVPRETAEEDGYLILFEVHDENTGKSCVTVIDAKTMSAPVAVELPHRV 518
QY 581 PYGFHGTFLIGADDLAKQV 599
DB 519 PYGFHALFYTEQDLQEQTL 537

RESULT 9
T51734
neoxanthin cleavage enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51734
R:Neill, S.J.; Burnett, E.C.; Desikan, R.; Hancock, J.T.
J. Exp. Bot. 49, 1893-1894, 1998
A:Title: Cloning of a wilt-responsive cDNA from an Arabidopsis thaliana suspension culture
A:Reference number: Z24454

[illegible]

C:Genetics:
A:Gene: all1106

Query Match 23.7%: Score 745.5; DB 2: Length 475;
Best Local Similarity 34.5%: Pred. No. 1.3e-48;
Matches 169; Conservative 91; Mismatches 171; Indels 59; Gaps 15;

```
OY 123 IAGNFAVNEQFVRRLPVVGLKLPDSIKGVYRNGANPLREPTGHHFPGDGMVAVKF 182
DB 22 LEGNFAVNEELTTDLKAYIGELPRLSGMFYRNGNPNPMTIPIGQHNHMDGDMHGVRI 81
OY 183 EHGASASYACRTQTNRFVQERQDGRPEPKAIGELHGTHTIARLMLFYRAAGIYD--- 239
DB 82 SNGKATYRRRYVTOFMQIHEHGAQAIW-----TGLMEPPQ 117
OY 240 ---PAHGTVANAAGLVYFNGRLAMSDDLPRYOQTTPNGDLKTVGRPDQGLESTMA 296
DB 118 TELPSRNTG--WTALIMHAGQILALMEGAPYAIQVP--DLASIGELYNNQDLSAFTA 172
OY 297 HKPVDSGELFALSYDVYSKYLFKFRSPDGTSPDVEIOLDDOPTMHDAITENFVY 356
DB 173 HKVDPVYTGEMMFEGS-FAPRYLHVSYSATGELVRYTPIIDLPKGVMMHDAITANTYI 231
OY 357 VPDQAVFKLPEMIRGSPVYVDKKNVAFGLDKYAEDSSNIKWIDARDCFCFLMNMW 416
DB 232 FMDLPLTFSEVRMQRGPMLMEFSDRSPRFGILPRHG-DNSQIRMFEDRSCYVHTLMAY 290
OY 417 EEPREDEVYVIG-----SCMTPPDSIFNESDENLK-SYLSERLNLKGTGSTRRIISN 469
DB 291 E--DDDEVLLFACPMKSTVILVSPS---QTPDEADIPRLHMRFLKLGKHEEML--- 342
OY 470 EDQOVNLEAGVNRNMLGKRTKFAVLA-LAE-PWPKVSGFAKVDLTGCVKKHLYDNRY 527
DB 343 --DDVASEPRRINENLTGPTQYGYTSRLAKGSIRPLFEBILIKDLSNASKSÖNEYOGKRY 400
OY 528 GGEPLFLEGESE-EDEGYILCFVHDEKTKWSELOIVNAVSEVE--ATVKLPSPRYVGE 584
DB 401 GSEAVFVPRPGATVEDDGWILTYVYDTGESESELVYINADINSEPIARILLPQRPVYGF 460
OY 585 HGTFIGADLL 594
DB 461 HGIVWTEQL 470
```

RESULT 11
T17019

hypothetical protein - apple tree
C:Species: Malus domestica (apple tree)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T17019
R:Matillon, B.; Kettmann, R.; Arredouani, A.; Hesquet, J.; Boxus, P.; Burny, A.
Plant Mol. Biol. 36, 909-915, 1998
A:Title: Apple messenger RNAs related to bacterial lignostilbene dioxygenase and plant S
A:Reference number: Z18655; MUID:98179104
A:Accession: T17019
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-446 <MAT>
A:Cross-references: EMBL:Z93765; NID:g2924324; PIDN:CAB07784.1; PID:g2924325
A:Experimental source: cv. McIntosh, strain Wjlcik

Query Match 23.7%: Score 745; DB 2: Length 446;
Best Local Similarity 39.8%: Pred. No. 1.3e-48;
Matches 150; Conservative 76; Mismatches 129; Indels 22; Gaps 9;

```
OY 231 ARAAGIYDPAHGTVANAGLVYFNGRLAMSDDLPRYOQTTPNGDLKTVGRPDQGLE 290
DB 76 ANVLGOVNPAPANGIGLANSLAFGDRYLALGESDLPYSLRLNSDITFLGHNDPKGL 135
OY 291 ESTMAHFKVDESGELFALSYDVYSKYLFKFRSPDGTSPDVEI-OLDDOPTMHNDPA 349
```

```
DB 136 SMMMTAHPKIDPDGTGEAFARFGFI-RPRLTYFRSDNSGVKQDPVFIISMVPTFLHDA 194
OY 350 ITENFVVPDQOVVFKLPFMI-RGSPVYVDKKNVAFGLDKYABDSSNIKWIDARDCFC 408
DB 195 ITKKAIFADIQIGLNLIDMTIKRATRFGLDPSKYVRIGVIRIYAKDEEMKRFVPGN 254
OY 409 CFHLMNAMEEPETDEVYVIGSCMTPPDSIFNESDENLKSVELIRLNLKTGSTRRPIIS 468
DB 255 GVHATNAWD--EDDALVMAVPNVLSAEHLYERVD-LVHCLVEKVRIDLKTGIYTRPIST 311
OY 469 NEDQOVNLEAGVNRNMLGKRTKFAVLAELFPRKVSFGAKVDLTGGE-----YKKNLYG 523
DB 312 R-----NDFEAVINPAVYLRKKNVYVAEGDPMPKISGVVKKLDVSNVEHKECTIVASRMG 366
OY 524 DNRYSGEPLFGE-----GGEDEGYILCFVHDEKTKWSELOIVNAVVS--LEAVATVKRP 577
DB 367 PGCYGGEFFVAREPENPAEDDNGFLVYVHDEKAGESRFLYMDAKSFQDLIVAAVRMP 426
OY 578 SRVPYGFHGTIGADLL 594
DB 427 RRPVYGFHGLFRESDL 443
```

RESULT 12
E87345

conserved hypothetical protein CC0776 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87345
R:Nietman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <STO>
A:Cross-references: GB:AE005673; NID:g13422015; PIDN:AAK22761.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0776

Query Match 16.8%: Score 528.5; DB 2: Length 483;
Best Local Similarity 30.9%: Pred. No. 4e-33;
Matches 160; Conservative 89; Mismatches 203; Indels 65; Gaps 22;

```
OY 108 EKLHPLPKTADPS--VOIAGNFAVNEQFVRRLPVV-GKLPDSIKGVYRNGANPLHBP 164
DB 4 ERLPVRTSLGPTNHPYMTGPWTPQHEEVANAMDLVLEGATPADDGVYLRNTENPHDP 63
OY 165 VVGHNEFFDDGMVHAVKFEHGSASYACRTQTNRFVQERQDGRPEPKAIGELHGTHTGIA 224
DB 64 IGRYHFFDDGMHIOIEFGGAATYRNRPRFRCFAEDAEVNEGIM---GGIMDGPVGS 119
OY 225 RLMLFYARAAGIYDPAHGT--GVANAGLVYFNGRLAMSDDLPRYO---VOITPNGDL 278
DB 120 K-----RPGFG---AHALNLDASATDLYVNGEALIAF-----YQCEATRLRDL-IL 163
OY 279 KTVGRFEDQGLESTIWAHPKVPDSGELFALSYDVYSK-PYLKYFRSPDGTSPDVE 336
DB 164 ENIGVASW-APLEG-VSAPRPVDEATGELMFNY---SKAMPYMHGVYVGPGRKKRYQG 218
OY 337 IOLDOPTMHNDPAITENFVVPDQOVVFKLPEMIRGSPVYVDKKNVAFGLDKYABDS 396
DB 219 VPLPGFRLPHDMAFSKRYAILNDLPVFMQDELMARIDIAHVAIRLHKGIPIRSFALVPR--EG 275
OY 397 SNIKWIDARDCFCFLHLMNAMEEPETDEVYVIGSCMTPP-----DSIFNESDEN 444
DB 276 GERPRFEAPRYVVLHNLNAYE--DGDEVYLDGYFOEKRPRLRELEGARPGHGLMAYLDH 333
OY 445 LKSVLSEI--RLNLKTGSTRRPIISNEDQOVNLEAGVNRNMLGKRTKFAVLAELP- 500
```

```

Db      334  --SFLPKLHMRMFNLKSTGETTTERKL-----DDRV-LEFGFNQKXAGKPRYRIVSTTARPG 386
Oy      501  WPKVSGFAKVDLTITGCEVKKHLGYDNRYGGEPLFLPDEGG-EEDEGYILCEVHDEKTMKSE 559
Db      387  WFLFNGFVKHDIETGESSIALPEGRYASEAFPAKVGAVDEDDGYLVSFFIDENKGASe 446
Oy      560  LQIVANVSLVEATVTKLPSSRVYGFHGFIGADDLAK 596
Db      447  CLIVDAKRFYVCRIALPHKLSSGTHYWAGNEMLETK 483

RESULT 13
A70534
hypothetical protein RV0654 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70534
R:CoLe, S.T.; BroSch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70534
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-501 <COL>
A:Cross-references: GB:295972; GB:AL123456; MID:93261790; PIDN:CAB09380.1; PID:e319166;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0654

```

```

Query Match Similarity      14.8%: Score 466; DB 2; Length 501;
Match Local Similarity     27.9%: Pred. No.2.3e-27;
Matches 148; Conservative 76; Mismatches 195; Indels 112; Gaps 18;

QY 123 IAGNEPYNEDQVRNRLPYVGKRLPDSIKGVYVRNANPLHE--PYTGHHFFEGDQKVHAY 180
      : | | | : : | | | : | | : | | | | | : : | | | | | | |
Db 14 LEGFLAPVSTEVATDLPVYTGRIPEHLDGRYLRNGPNPAVEVDPAT-YIMFPGDAMVHG 72

QY 181 KPEHSSASACFEYQTNRFVQROJGRVPFPAIGE-----LHGHTGIALMLFYARAAA 235
      : : : : : | : : : : | | | | | | | | | | | | |
Db 73 ALRDKA-----KMYRRRWRTPAVCALGEIPISARHPRTGILE----- 112

QY 236 GIVDEAHGTVYANAGLIVYFNGRLLMSEEDDLRYOQIIFPNGDKLTYGRFEGOLESTWI 295
      : | : : : | | : : | | | | : | | | | | | | | | |
Db 113 -----GGNTNVLTTHAGRTLALVEGV---VNYELTDELDVYQCPDFEGTILHGCT 160

QY 296 AHPKYDPESGELFALSYDYVSKPYLKYREFESDGTSPDEYIOLDQPTMHDAITENFV 355
      : | | | : | | | | : : | | : | | : | | : | | : | |
Db 161 AHPDDPHTGELHAAYSSEFARGHRYQSVITGDGHARRVDIEVAGSPMMHSEFSLTINY 220

QY 356 VVPDOQVVE-----KLPMTI-----RGGS---PV 376
      : | : | : | : | : | : | : | : | : | : | : | : |
Db 221 VYIDLPYTFDPMQVVPASVPRWLQRPARYLIQSYLGRVYRIPDIALALGRMGHSDRLXY 280

QY 377 VYDKNRVAFGLIDKYAEDSSNIKWIDAPDCFCFHLNAMEEPETD-EYVYVG---SCM 431
      : | | | : | | : : : : | | : | | : | | : | | : | |
Db 281 AWPSPYPAVGVMPREG-GNEDYRMFDEIPCYVYHPLNYSRCNGAEVLVDVYRYSRM 339

QY 432 TPDPDSIFNSEDNLKSVLSEIRLNKLTGSGTSPRIISNEDQVNLDEAGVNRNMIGRTK 491
      : | | : | | : | | : | | : | | : | | : | | : | |
Db 340 EDRDRRGPGGDS--RPISDRMTINLATGAVTAE--CRVDRA--OEPRININTEVLGPPHR 392

QY 492 FAY-----LALAEPMYPVSGFAKVDLTGTGEVKHMLGDNRRGGEPFLFPGEG 538
      : | | : | | : | | : | | : | | : | | : | | : | |
Db 393 FAYTVGIEGGLVLVGAALSTP-----LYKQDCVYTGSSIVASLDPDLIGEMVFPNRS 446

QY 539 GE-EDBGYILCYVHDEKTKSELQYLVNNAVSELEAVETKLPISRVYFGHGT 588
      : | | : : : : : : : : | | | | | | | | | | :

```

DB 447 AAAEDDGLMGVGMWRGRDEGQLLLDAQTLSEIATVHLPQKRVPMFGNGM 497

RESULT 14

A70582

hypothetical protein RV0913c - *Mycobacterium tuberculosis* (strain H37Rv)

C:Species: *Mycobacterium tuberculosis*

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70582

R:COLE, S.T.; BROSCH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON; CONNOR, R.; DAVIES, R.; DEVLIN, K.; FELTWELL, T.; GENTLES, S.; HAMLIN, N.; HOLLOYD; RANDJADREEM, M.A.; ROGERS, J.; RUTTER, S.; SEEGER, K.; SKELTON, S.; SQUARES, S. Nature 393, 537-544, 1998

A:Authors : Squares, R., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A:Reference number: A70500; MUID:98295987

A:Accession: A70582

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-502 <COL>

A:Cross-references: GB:295210; GB:AL123456; NID:g3261757; PIDN:CAB08511.1; PID:e3152125

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV0913c

Query Match	Similarity	14.4%	Score 454.5	DB2	length 502
Best Local	Similarity 27.6%	Pred. No. 1.7e-26			
Matches 139	Conservative 73	Mismatches 190	Indels 101	Gaps	
QY	141	VVGKLPDSIKGVVVRNCGANPLHEPVTGHHFFEDDGGHVAHVKKPEHGSASYACRFETQTRNFV	200		
		: : : : : : : : : : : : :			
Db	41	VVGVEPADLDGILYLRNTENPLHPAFATYHPDFDGDGMHVFGFRDKAFYRNFRIFRTDGL	100		
QY	201	QEROLGRVFFPKALGELGHGTARLMLFYAR---AAAGYDPAIGTGV-ANALV---	252		
		: : : : : : : : : : : : : : : : : :			
Db	101	AENEGGLWP-----GLAEPLQAKRHHGAGRLMKMDASSTDVIVHRIALTLS	150		
QY	253	-YFNRLTAMSEDDLPYQVOITPNGDLKTVGREFDGGOLESTMIAPHKVDPESELFALS	311		
		: : : : : : : : : : : : : : : :			
Db	151	FVGGDDLYRID---PYSANTL--GKESMHGRRFPFGVGS---AHPKVDNKTGGLLFPN	200		
QY	312	YDVYSKPLKLFRRSPDGTSPDYEIOLDPPTMHHPAITEFNFYVYPDQVYFKLPEMR	371		
		: : : : : : : : : : : : : : : :			
Db	201	YS-KOEPRMRGVYVDONNELVHYVDVLPRLPRHDMAFENYVILND-----	247		
QY	372	GSPVYVYKKNVYARFGLIDKY-----AEDSSNIKIMDAPDPCFCFHLNMEPE	420		
		: : : : : : : : : : : : : : : :			
Db	248	-FLEMDPRLLENDYHLPRYPIPRFAVVARRGNDIRFPAADPTFLHTNMYE--Q	303		
QY	421	TDEVVYVIGSCMTPPDSIFNESP-----ENLKSVLSEIRLMKTGE	460		
		: : : : : : : : : : : : : : : :			
Db	304	GGEIYLDG-----FYEEDPQPLDTGSGTKWEKLFRLALDRLOSRIHRRRLMWVG-	353		
QY	461	STRRLTIENEDQVY---LEAGVNRNMLGRKTKFPYVILAAE-WPKYSGFAYDLTJGE	516		
		: : : : : : : : : : : : : : : :			
Db	354	-----AYHEOQLSESTTEGTTINADYASSTRYYAAATGKSWFLFDGLYKHDLTGN	406		
		: : : : : : : : : : : : : : : :			
QY	517	VKKHLGDNRYGEBPLPLPBGGE-EDEGYILCFVHDEKTKWSELOIVNAVSLVEAYTK	575		
		: : : : : : : : : : : : : : : :			
Db	407	HECYSEFGDGYVGSFTAMAPRVGSSAEDDGYLVLTDDMDNDASVCLVFPDAPRGDPICK	466		
QY	576	--LBSRVYFGHGFICADDIAK	596		
		: : : : : : : : : : : : : : : :			
Db	467	IALPERISSGTHSAMVGAELRR	489		
RESULT	15				
	876169				
	hypothetical protein - <i>Synechocystis</i> sp. (strain PCC 6803)				
	C:Species: <i>Synechocystis</i> sp.				
	C:Variate: PCC 6803				
	C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999				

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2002, 10:12:13 ; Search time 13.49 Seconds

(without alignments)
1719.274 Million cell updates/sec

Title: US-09-758-269-6

Perfect score: 3150
Sequence: 1 MASFTATAVSGRWLGNGNHT.....VPYGHGFTIGADDLAKOV 599

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114	3.6	870	YC50_YEAST	P25623 saccharomyc
2	113	3.6	466	DNAA_PROMI	P22837 proteus mix
3	107.5	3.4	658	PAKL_SCHPO	P50527 schizosacch
4	106	3.4	3562	PGCV_CHICK	O90953 gallus gall
5	105	3.3	783	CAD5_MOUSE	P55784 mus musculi
6	105	3.3	2504	FAS_HUMAN	P49927 homo sapien
7	104.5	3.3	1079	APCE_PREDI	P16566 tremyella d
8	103	3.3	775	SYPB_AQUAE	O67620 aquilae aeo
9	102	3.2	2493	CYAA_USTMA	P49606 ustilago ma
10	101.5	3.2	520	GLGS_BRANA	O9m462 brassica na
11	101.5	3.2	565	SCRL_SCHPO	O14335 schizosacch
12	101	3.2	512	GLGT_VICFA	P52417 vicia faba
13	100	3.2	3770	ACVS_EMENT	P27742 emeticella
14	99.5	3.2	383	PHYC_BACSU	O31097 bacillus su
15	99.5	3.2	1256	ATL_STRAU	P52081 staphylococ
16	98	3.1	741	MASZ_MYCTU	O09596 mycobacteri
17	98	3.1	1788	VP72_CAERE	O09921 caenorhabd
18	97.5	3.1	518	CP74_ARATH	O09642 arabidopsis
19	97.5	3.1	691	SLP1_YEAST	P20795 saccharomyc
20	97	3.1	636	KDGP_SCHPO	O14019 schizosacch
21	96.5	3.1	426	UCR2_SCHPO	P78761 schizosacch
22	96.5	3.1	503	DLFA_BACSU	P39581 bacillus su
23	96.5	3.1	862	SMR2_YEAST	P32567 saccharomyc
24	96.5	3.1	3587	TYCB_BACBR	O30408 b tyrocidin
25	95.5	3.0	865	NRAA_PENUR	O92269 penicillium
26	95.5	3.0	918	DNLI_RAT	O9jny8 rattus norv
27	95.5	3.0	1144	DP3A_NEIMB	O29555 neisseria m
28	94.5	3.0	590	HMAA_DROME	P29655 drosophila
29	94.5	3.0	638	GHR_RAT	P16310 rattus norv
30	94.5	3.0	953	LKAB_PASHA	P55118 pasteurella
31	94.5	3.0	1046	LPDC_WEITHE	P96177 weissella h
32	94	3.0	532	P72_MYCMY	P55801 mycoplasma
33	94	3.0	919	Y893_HUMAN	O94967 homo sapien

34	93.5	3.0	1275	GNRP_HUMAN	Q13972 homo sapien
35	93	3.0	3067	CAIC_MOUSE	O60847 mus musculi
36	92.5	2.9	354	OTX1_HUMAN	P32242 homo sapien
37	92.5	2.9	955	LKAA_PASHA	P55117 pasteurella
38	92.5	2.9	1057	SPSL_CITUN	O22060 citrus unsh
39	92.5	2.9	1116	SLPH_BACBR	P38538 bacillus br
40	92.5	2.9	1131	CHSL_YEAST	P08004 saccharomyc
41	92.5	2.9	4829	BIR6_HUMAN	O9nr09 homo sapien
42	92	2.9	880	CADP_XENLA	P33148 xenopus lae
43	91.5	2.9	474	GSAL_ARATH	P42799 arabidopsis
44	91.5	2.9	744	HXCI_HAEIN	P44523 haemophilus
45	91.5	2.9	896	APCE_SYNY4	Q02907 synechocyst

ALIGNMENTS

RESULT 1	YC50_YEAST	STANDARD:	PRT:	870 AA.
ID	YC50_YEAST	P25622:		
AC	01-MAY-1992 (Rel. 22, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
GN	Hypothetical 96.1 kDa protein in RIM1-PSI4A intergenic region.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OX	Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.			
RN	NCBI_TaxID=4932;			
RP	SEQUENCE FROM N.A.			
RA	Cederberg H., Hohmann S., Schaaff-gerstenschlaeger I., Huse K.,			
RL	Zimmermann F.K.,			
RM	Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.			
RP	REVISIONS.			
RA	Gromadka R.;			
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.			
CC	- SIMILARITY: SOME, TO S.POMBE SPBC4C3.06.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL: X59720; CNA42322.1; -			
DR	SGD: S0000626; YCR030C.			
KW	Hypothetical protein.			
SQ	SEQUENCE 870 AA: 96125 MW; 3F2CE2P1562E5277 CRC64;			
Query Match	3.6%; Score 114; DB 1; Length 870;			
Best Local Similarity	20.6%; Pred. No. 1.2; 195; Indels 180; Gaps 37;			
Matches 121; Conservative 92; Mismatches 180; Gaps 37;				
QY	3 SFPTAAVSGRWLGNGNHTQPLSSSOSDLS---YCSLPMASRYTKLWVSAL--HTP 57			
DB	384 SSTATSIYK-----QRRRTYSSSKSNMWTPEASDTPPLPHATPK-NVDAFVADTP 437			
QY	58 PALHF-----PKSSNSPAIVVKKAKESNTKOMNLFQRAAAALDAAGFLVSHE 108			
DB	438 PAQTFPPESEVPSTPOSS-----PPTAKEDPSNL-----PKTVPIISIPLOPOS 485			
QY	109 KLHPLF-KTADPEV-----QIAGNFAPVNEOPRRRLPVVGKLPDSIKGVYRNGA 158			
DB	486 KTKPLPLEPSPSISLPTATVNOPSGQ---VSRPLHTRAPA---LPSSRKONFTIHRD 539			
QY	159 NPLHEPVTHGHFFDGDGMVAHVAFEHGASAYACFTQTNRFVDEKRGREPV---FFKAIG 215			


```

Db 540 SOLYSLIP-----NHGSGA-----TPTSSLSLIPDERPVSTLSSQITG 578
QY 216 ELHGHTGIALRLMEFYARAAGIVDP-AHGTVANAGLVYFNGRLAMSEDDLPYOVITP 274
Db 579 ELKE-----LNPKATGSSSTISVSGSLFQHSPLDTSQGLNASTAEVL 620
QY 275 NGDLKTVGRPDFDQGLESTMIAPHKYDPESELFALSY---DVVSKPYLKYFRFSPDGTK 331
Db 621 NASEFK-----DGLMNSQOLI-----GEI-ALNYLPSNVNSPL-----PIG-- 655
QY 332 SPVEVQLDQPTMHDPATLENFV---VPPDQVVFKEPLKEMIRGSPVYVDKKNVARGCIL 389
Db 656 ---INLRINGAKFEKYLINQAFIERVAPPE---FKV-----NPSFIDSRTLGAI--- 699
QY 390 DKYAESSNIKWIDAPDPCFCFH-LMNAMPEPETDEVVIGSCMTPPSI----- 437
Db 700 -KI-----SIKEPIAP--IYIHPPWR-FESHQASVVLTVKMSPSLPBEISQIYIEDLVVF 750
QY 438 FNESDENLKSVLSEIRLINIKTGEST---RPII---SNEDQV---NLEAGNVNRM 486
Db 751 VNIDGAMATGALSQPGSFKSKKKRITWRFEKPVLTFRNEGQRLIARFTDGLAHBSAK 810
QY 487 GRRTKFAVYALAE---PMPKVSQFAVVDLTGGEVKKHLXGDNRYGGE 530
Db 811 GVITKF---TISETDNALPHSGSGSGLTTCQELDE---NNPFGE 851

RESULT 2
DNA_PROMI
ID DNA_PROMI STANDARD; PRT; 466 AA.
AC P22837;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chromosomal replication initiator protein dnaA.
GN DNA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LAI509;
RX MEDLINE=91033012; PubMed=2172087;
RA Skovgaard O.;
RT "Nucleotide sequence of a Proteus mirabilis DNA fragment homologous
RT to the 60K-rnpA-rpmH-dnaA-dnaN-recF-gyrB region of Escherichia
RT coli."
RL Gene 93:27-34(1990).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE INITIATION AND REGULATION
CC OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN OF REPLICATION;
CC IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA AT A 9 BP CONSENSUS
CC (DNA BOX): 5'-TTATTC(C/A)A(C/A)A-3'. DNAA BINDS TO ATP AND TO
CC ACIDIC PHOSPHOLIPIDS.
CC -1- SIMILARITY: BELONGS TO THE DNAA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: M58352; AAA83958.1; -
DR PIR: J00733; IOEBV.
DR InterPro: IPR001957; Bac_DnaA.
DR Pfam: PF00308; bac_dnaa; 1.
DR PRINTS: PR00051; DNAA.
DR PROSITE: PS01008; DNAA; 1.
KW DNA replication; DNA-binding; ATP-binding.
FT NP_BIND 171 178
SQ SEQUENCE 466 AA; 52974 MW; F8B67C142FE9FA41 CRC64;

```

```

Query Match 3.6%; Score 113; DB 1; Length 466;
Best Local Similarity 21.9%; Pred. No. 0.56;
Matches 84; Conservative 48; Mismatches 112; Indels 140; Gaps 21;

QY 105 VSEHKLHPLPKT-ADPSYIAGNFAVPYNEQPVRRNLFPVGLKLPDS-IKGYVRGANGPLH 162
Db 83 VSARTTESVPEKTYTHPAV-----NSTPTNSQPVK---PSMDNQPSQLPELNTSRNVNPKH 135
QY 163 EPTVGHNFEPDGDGVNNAVKFEHGSASACFTQTNRVQEROLG--RPVYPKAIGELHGH 220
Db 136 K-----FD-----NVEGKSQOLAR-AAARQVANPGGAVNPLF-----LYGG 172
QY 221 TGIARLMLFYARAAGIVDPAHGVY---ANAGLVYFNGRLAMSEDDLPYOVITPNG 276
Db 173 TGLGKTLHLH-----AVGNSIYERRKANAKVYVMH-----SRFQVDYKALQNN 216
QY 277 DLKTVGRPDFDQGLESTMIAPHKYDPESELFALSYDVVSKPYLKYFRFSPDGTKSPDVE 336
Db 217 AIE-----DFK-----RYYR-SVDALLIDDIQ 237
QY 337 IQLDQPTMHDPATLENFVYVPPDQVVF---KLPEMIRGSPVYVDKKNVARGCILDKYA 393
Db 238 FFAKKERSQEEFHTFVALLEGNOQILITSDRPKEING---VEDRLK-SRFG----- 286
QY 394 EDSNIRKIDAPDPCFCFHLNMNAMESPETDEVVIGSCMTPPSIFNESDENLKSVLSEI- 452
Db 287 -----WGLVVAIEPPELETFRVAL-----LMKRADENOIOLPDEVA 321
QY 453 -----RLNLKGTGSTRPPIISN 469
Db 322 FFIKKRLRSNVRELEGALNRYVIAN 345

RESULT 3
PAK1_SCHPO
ID PAK1_SCHPO STANDARD; PRT; 658 AA.
AC P50527;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Serine/threonine-protein kinase pakt1/shk1 (EC 2.7.1.-).
GN PAK1 OR SHK1 OR ORB2 OR SPBC1604.14C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96112805; PubMed=846783;
RA Ohtllie S., Miller P.J., Johnson D.I., Creasy C.L., Sells M.A.,
RA Bagrodia S., Forsburg S.L., Chernoff J.;
RT "Fission yeast pakt1 encodes a protein kinase that interacts with
RT Cdc42p and is involved in the control of cell polarity and mating."
RL EMBO J. 14:5908-5919(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Marcus S.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Beck A., Reinhardt R., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 119-658 FROM N.A.
RX MEDLINE=95320235; PubMed=7597098;
RA Marcus S., Polymerio A., Chang E., Robbins D., Cobb M.H.,
RA Wiggler M.;
RT "Shk1, a homolog of the Saccharomyces cerevisiae Ste20 and mammalian
RT p55PAK protein kinases, is a component of a Ras/Cdc42 signalling

```


CC	long-chain fatty acid + (N+1) COA + N CO(2) + 2N NADP(+).
CC	-1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].
CC	-1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].
CC	-1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) + [acyl-carrier protein].
CC	-1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC	-1- CATALYTIC ACTIVITY: (3R)-3-hydroxypantimtoyl-[acyl-carrier protein] = 2-hexadecenoyl-[acyl-carrier protein] + H(2O).
CC	-1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC	-1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2O) = [acyl-carrier protein] + oleate.
CC	-1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC	-1- TISSUE SPECIFICITY: PROMINENT EXPRESSION IN BRAIN, LUNG, AND LIVER.
CC	-1- MISCELLANEOUS: THE RELATIVELY LOW BETA-KETOACYL SYNTHASE ACTIVITY MAY BE ATTRIBUTABLE TO THE LOW 4'-PHOSPHOPANTHETHEINE CONTENT OF THE PROTEIN.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sdb.ch/announce/or_send_an_email_to_license@sib-sdb.ch).
CC	--
DR	EMBL; U26644; AAC50259.1; --
DR	MIM; 600212; --
DR	Interpro: IPR001227; Acyltransf.domain.
DR	Interpro: IPR002085; Ach_zn_family.
DR	Interpro: IPR000794; Ketoacyl-synt.
DR	Interpro: IPR003880; Phosphopant_attach.
DR	Interpro: IPR001051; SAM_bind.
DR	Interpro: IPR001031; Thioesterase.
DR	pfam: PF006968; ACYL_transf. 1.
DR	pfam: PF00107; adh_zinc. 1.
DR	pfam: PF00109; ketoacyl-synt. 1.
DR	pfam: PF02801; ketoacyl-synt_C. 1.
DR	pfam: PF00550; pp-binding. 1.
DR	pfam: PF00975; Thioesterase. 1.
DR	Pfam: PS00606; B_KETOACYL_SYNTHASE. 1.
DR	Prosite: PS00606; B_KETOACYL_SYNTHASE. 1.
DR	Prosite: PS0075; ACP_DOMAIN. 1.
KM	Fatty acid biosynthesis: Multifunctional enzyme: Phosphopantetheine; Hydroxylase; Oxidoreductase; Transferase; Lyase; NADP; Pyridoxal phosphate.
FT	DOMAIN 1 413
FT	DOMAIN 428 815
FT	DOMAIN 1630 1857
FT	DOMAIN 1858 2113
FT	DOMAIN 2118 2174
FT	DOMAIN 2202 2504
FT	ACT_SITE 161 161
FT	ACT_SITE 580 580
FT	ACT_SITE 876 876
FT	NP_BIND 1666 1683
FT	BINDING 1689 1699
FT	NP_BIND 1880 1895
FT	BINDING 2151 2151
FT	ACT_SITE 2302 2302
FT	ACT_SITE 2475 2475
SQ	SEQUENCE 2504 AA; 273100 MW; 8AAP9809B2338DFA CRC64;

```

OY 11 SGRRLNGHNTQPELSSQSSDLSYSSSLPMASRVTRKLN--VSSALHTPPALHPFKOS-- 66
Db 707 SARWL-----STIPEAQMHSSLARTSAEYVNNVLSPVLFQEDALMHPERHVV 756
OY 67 -----SNSPAIVYKKKAKESNTKQKNLLEFORAAAAALDAEGFVSHKRLHP,PKTADPSV 121
Db 757 LEIATPPCRQAVLKKVRKPSCT---ILPRMKKHRLDLEFPLAGIGLHLTSGIDANANA 812
OY 122 QIAGNFAPVNEQPVRRNLVYVGKLPDSIKGYVVRGANGPLHEPVTHGHFPDGDGVAAVK 181
Db 813 L-----FPPV-ESPADRGPIPLISPL-----IKMHSLSMWADAPA-AED 847
OY 182 FENSGAS-----YAC--RFTQTRNVQERQL-GRVFPRKATIGELGHGIRMLFYARA 233
Db 848 FPNSGGSPSATIYTTPTPSESESPDRYLVDHTIDGRVLFPAT---GYLSIWKTL--ARA 900
OY 234 AAGI-----VPAHGTGVANGVLYFNGRLLAMSEDDLPYQVQITTPNGDLKTGVR- 283
Db 901 WAGLEQLPVFEVDVYQHQTILPKTGYVSLERLL-----EATGAFEVSENGNLVYSGKV 955
OY 284 FDFDQLESTMLAHPR-VDPES--GELPALSVDVSKPELYKFRSP 327
Db 956 YQWMDP-DPLRFDHPESPHPNSPRSLPLAQAEVYKELRLNGYDGP 1001

RESULT 7
APCE_FREDI STANDARD: PRT; 1079 AA.
AC PI5566:
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Phycobllisome 120 kDa linker polypeptide, core (L-CW 92) (Core-
DE membrane linker protein).
DE APCE.
OS Fremyella diplosiphon (Calothrix PCC 7601).
OC Bacteria; Cyanobacteria; Nostocales; Rivulariaceae; Fremyella.
OX NCBI_TaxID=1197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90192765; PubMed=2107546;
RA Hounard J., Capuano V., Colombano M.V., Courdin T., de Marsac N.;
RT "Molecular characterization of the terminal energy acceptor of
RL cyanobacterial phycobilisomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2152-2156(1990).
CC -1- FUNCTION: THIS PROTEIN IS POSTULATED TO ACT BOTH AS TERMINAL
CC ENERGY ACCEPTOR (BY ITS PHYCOBILLIN-LIKE DOMAINS) AND AS A LINKER
CC POLYPEPTIDE (BY ITS REPEATS AND ARMS) THAT STABILIZES THE
CC PHYCOBILLISOME CORE ARCHITECTURE.
CC -1- SUBUNIT: PHYCOBILLISOMES OF THIS ORGANISM ARE COMPOSED OF A TWO
CC CYLINDER CORE, FROM WHICH SIX RODS RADIATE. THE CORE IS MAINLY
CC COMPOSED OF ALLOPHYCOCYANIN ALPHA AND BETA CHAINS, AND OF THREE
CC MINOR COMPONENTS: THE ALLOPHYCOCYANIN ALPHA-B CHAIN, A 18.3 kDa
CC POLYPEPTIDE, AND THE ANCHOR POLYPEPTIDE LCM.
CC -1- SUBCELLULAR LOCATION: ANCHORS THE PHYCOBILLISOME PERPENDICULARLY
CC TO THE STROMAL SURFACE OF THE THYLAKOID MEMBRANE.
CC -1- SIMILARITY: THE REPEATED DOMAINS ARE SIMILAR TO THE N-TERMINAL
CC REGIONS OF PHYCOCYANIN ROD LINKER POLYPEPTIDES.
CC -1- SIMILARITY: THE PHYCOBILLIN-LIKE DOMAINS ARE SIMILAR TO PHYCOBILLINS
CC FROM VARIOUS SPECIES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20806; AAA24873.1; -.
DR PIR: A35088; A35088.
DR HSSP: P00318; IBS3.
DR InterPro: IPR001297; PBS_linker-poly.

```

DR InterPro: IPR001659; Phycobillosome.
DR Pfam: PF00427; PBS_Linker_poli; 4.
DR Pfam: PF00502; Phycobillosome; 2.
KM Phycobillosome; Electron transport; Photosynthesis; Repeat.
FT INIT_MET 0
FT DOMAIN 17 75 PHYCOBILIN-LIKE 1.
FT DOMAIN 76 143 PHYCOBILIN-LIKE LOOP.
FT DOMAIN 144 236 PHYCOBILIN-LIKE 2.
FT DOMAIN 237 284 ARM 1 (SPACING SEQUENCE).
FT REPEAT 285 409 I.
FT DOMAIN 410 546 ARM 2 (SPACING SEQUENCE).
FT REPEAT 547 669 II.
FT DOMAIN 670 743 ARM 3 (SPACING SEQUENCE).
FT REPEAT 744 869 III.
FT DOMAIN 870 953 ARM 4 (SPACING SEQUENCE).
FT REPEAT 954 1079 IV.
SQ SEQUENCE 1079 AA; 120325 MW; 87FE38F232BFC82 CRC64;

Query Match 3.3%; Score 104.5; DB 1; Length 1079;
Best Local Similarity 21.3%; Pred. No. 8.1;
Matches 71; Conservative 47; Mismatches 80; Indels 135; Gaps 21;

OY 62 PFKQSSN---SPAIVYKPKAKESNTKONLFORAALDAEGFLVSEKILHP-----L 113
DB 467 FKEKTRNPTSPA---PFSKDRIRILN-----QGPGINSQVSNPGARGEF 509
OY 114 PRTADPSV---QIAG-----NEAPVNEQPYARNLPVYGKLPDSIKGYVVRNG 157
DB 510 PGLSLPKVFRILDQLPDTIGTKAKAKASIKFESSSTQAV-----IKAAVLQ-- 554
OY 158 ANPLHEPTYGHHFFDGDGM-VHAKVEHGSAS-----YACREFT 194
DB 555 -----VGRGVYVGGQRLKVGKLENGQLSVREFIRALAKSDVFRKTYWTSLVYCKAI 607
OY 135 QTNREVOERQLGRPVYKPAIGELHGTGIALMLFYARAAGIVDPAGHTGVANAGLYE 254
DB 608 E---YIHRRLGLRPYVGRG--EINKYFDIAAKQGFYA-----VVD---AIIINS--VEY 650
OY 255 NGRLLAMSEDDLPYVOVITPNNG---DKTVG--RFDPGQLESTMIAHPKYDPEGELF 308
DB 651 SE---AFGEDVYVEREYLLPSSGVALLRQLR-VGSIREVDVG-----KVQKQETPLF 696
OY 309 ALSYDVNSKPYLKYFRFSPDGKTSPEVDIEQLDQ 341
DB 697 VTLGTVT-----DTRTEPDIOGRING 717

RESULT 8
SYFB_AQUAE STANDARD; PRT; 775 AA.
AC 067620;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
GN tRNA ligase beta chain) (PheRS).
OS PHET OR AQ.1730.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OC NCBI_TaxID=63363;
OX
RN
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RC MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.J., Keller R.,
RA Feldman R.A., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
aeolicus";
RL Nature 392:353-358(1998).
CC -I- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).

CC -I- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC SIMILARITY).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE PHENYLALANYL-tRNA SYNTHETASE BETA
CC CHAIN FAMILY. SUBFAMILY 1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AE000753; AAC07582.1; -
DR HSRP: P27002; 1PYS.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KM Complete proteome.
SQ SEQUENCE 775 AA; 88484 MW; 61BB15FDE5B274C8 CRC64;

Query Match 3.3%; Score 103; DB 1; Length 775;
Best Local Similarity 20.7%; Pred. No. 6.5;
Matches 114; Conservative 67; Mismatches 159; Indels 212; Gaps 31;

OY 123 IAGNEFAPVNEQPYARNLPVYGKLPDSIKGVYVRNGANPLHEPTGHHFDP----- 172
DB 46 VFGKVEVGEHHPYKRLAVY-----KVQYOHIFIDVYTVDKSVRE 86
OY 173 GGMVHAKVEHGSASACRFTQTNREVOERQLGRPVYKPAIGELHGTGIALMLFYAR 232
DB 87 GDGVI--VALPNAKVGNMC-----VTEREF--DGVSQGLLSA 121
OY 233 AAGIYDPAHGTGVANAGLYVRNGRLAMSEDDLP-----YQVQITPN-GD 277
DB 122 QELGLEEKSEG-----VLKIHEDFKPDTANEILGEGEKIIEIDITPNRQD 167
OY 278 LKTVGFRFDQGLESTMIHPKVPD-----ESGELFALSYDVNSKPYLKYFRFSPDGK 331
DB 168 MLSVGRVARD--LSAIFLPKPKKPEPPEYETGEFFIEIDEDCKRY-----RGVY 216
OY 332 SPDVETQLDQP-----TMMHDAFITEENFVYVDQGVFKLPIMRIGSPV-VY 378
DB 217 IGVGEIK--ESPLYIKRLMQCGIKSINNVDITNVY-----MLRDQGLHAF 262
OY 379 DNKVARFGILDKYADSSNIMKIDAPDCFCFLNMANMEPEPTDEVYVIGSCMTPPDSIF 438
DB 263 DLSKV-EGGIIVRSAKKGEKIITLDG-----EEREIDEDILVIADREKRLAVA 309
OY 439 N-----ES--DENLKSVLSEI-----RLNLKTGSTRRPIISNEDQVNLBA 478
DB 310 GVIIGLESIKENTKDILLESAYFNFPYRKASKKGIOTESSYR-----FERNVDIR- 362
OY 479 GAVNRNMLGRKTKFA-YIALAPBMPKVGSAKAVDLTTGEGKHLXGDNRYGGEPLFLP-- 535
DB 363 -----RVDRADQYAVYLLIKHAGKV-----KVVAQDYRE-KYRKPKVYLLPG 404
OY 536 -----GEGGEDGECITLCFVHDEKTKWKSQDIIVANVSLFENATYKLPBSRVYRGFGT- 588
DB 405 KYIRVAGESYKNNE-----VKEIIDALEIPMEI-MRCGVLEVLP3-----HRSFD 448
OY 589 IGAD-DLAKQV 599
DB 449 IQRDQDLIEIM 460

RESULT 9
CYAA_USTMA STANDARD; PRT; 2493 AA.
AC P49606;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenyl1
DE cyclase).
GN UAC1 OR REML.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_Taxid=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=518.
RX MEDLINE=95087882; PubMed7995519;
RA Gold S., Duncan G., Barrett K., Kronstad J.W.;
RT "CAMP regulates morphogenesis in the fungal pathogen Ustilago
RT maydis.";
RL Genes Dev. 8:2805-2816(1994).
CC -!- FUNCTION: PLAYS ESSENTIAL ROLES IN REGULATION OF CELLULAR
CC METABOLISM BY CATALYSING THE SYNTHESIS OF A SECOND MESSENGER,
CC CAMP.
CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -!- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-3 FAMILY.
CC -!- SIMILARITY: CONTAINS 21 LEUCINE-RICH REPEATS (LRR).
CC -!- SIMILARITY: CONTAINS A PP2C-LIKE DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL, L33918; AAA57469.1; -!
DR InterPro: IPR001054; Guanylt_cyclase.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_cyp.
DR InterPro: IPR001932; PP2C_domain.
DR Pfam: PF00211; guanylate_cyc; 1.
DR Pfam: PF00560; LRR; 18.
DR Pfam: PF00481; PP2C; 1.
DR SMART: SM00044; CYCC; 1.
DR SMART: SM00370; LRR; 6.
DR SMART: SM00369; LRR_Typ; 2.
DR SMART: SM00332; PP2C; 1.
DR SMART: SM00314; RA; 1.
DR PROSITE: PS00125; GUANYLATE CYCLASES_2; 1.
KM Lyase; Repeat; Leucine-rich repeat; CAMP synthesis; Magnesium.
FT DOMAIN 759 763
FT DOMAIN 888 897
FT REPEAT 1084 1106 LRR 1.
FT REPEAT 1108 1131 LRR 2.
FT REPEAT 1132 1155 LRR 3.
FT REPEAT 1157 1178 LRR 4.
FT REPEAT 1179 1201 LRR 5.
FT REPEAT 1202 1225 LRR 6.
FT REPEAT 1227 1248 LRR 7.
FT REPEAT 1249 1271 LRR 8.
FT REPEAT 1273 1295 LRR 9.
FT REPEAT 1314 1337 LRR 10.
FT REPEAT 1338 1360 LRR 11.
FT REPEAT 1361 1384 LRR 12.
FT REPEAT 1386 1407 LRR 13.
FT REPEAT 1408 1430 LRR 14.
FT REPEAT 1432 1455 LRR 15.
FT REPEAT 1509 1529 LRR 16.
FT REPEAT 1533 1556 LRR 17.
FT REPEAT 1557 1580 LRR 18.
FT REPEAT 1581 1604 LRR 19.
FT REPEAT 1606 1628 LRR 20.
FT REPEAT 1633 1656 LRR 21.
FT DOMAIN 1722 2001 PP2C-LIKE.
FT DOMAIN 2002 2493 CATALYTIC.

SQ SEQUENCE 2493 AA; 271979 MW; 106A872C3C1C5BF8 CRC64;
Query Match 3.2%; Score 102; DB 1; Length 2493;
Best Local Similarity 21.7%; Pred. NO. 42;
Matches 85; Conservative 44; Mismatches 136; Indels 126; Gaps 17;
QY 4 FTATAVSGRWLGNNH-----OPPLSSOSSDLSYCSLLPMASVYTKLVNVSALHTP 57
DB 801 FKKTSLAGSRGCTDSDVPLTALPLPGSKSVDEAAANKVYLDQ-TNNLQSAALYVO- 858
QY 58 PALHPPQSSNSPDAIVKPKAKESNTKQNNLFORAAAAALDAEGLVSH---EKLHPL 113
DB 859 ---QSOQHHPSPVRYRTSRGAGAHW--PASAGASAAAAGKLGHRPSKRHMNR 913
QY 114 PRTAD-----PSYQAGFAFVNEQPVARNLPYVGKLPDSIKGYVVRGAPLHEPTG 167
DB 914 PNTAGSVGATRPSTYTLTGSLSAEDD-----TSINGSIKRDG-HPLRSATA 959
QY 168 -----HHFFDGDGMVAHVFEHGSASVACRFQTQNNFVQERQGRVFPKAIGE 216
DB 960 NNNMATGTLPRNH-----IKYKTDGTFATLSCSVSTANEVQ----- 998
QY 217 LHGHTGIARLMDFYRAAAGIVDPAHGTGVANAGLVYENGRLAMSEDDLPRQVQ----- 271
DB 999 ---TLARKSLTTSAAVRLF-----VRDKG---SERDLGIS--DKPSQLRRRLI 1041
QY 272 ---TTPNGDLKTVGRFDPFGQLESTMIAPKVPDESGELFALSYDVSKPYIKYRFSFD 328
DB 1042 QAGYENDLEDEMGRDLS-----YLLREVFVRD 1070
QY 329 GTRKSPDVE-IQDOPMMDFAITENFVVVP 358
DB 1071 SVPTFDSISIGSEHTFOHLDHSRLNEMVP 1101
RESULT 10
GLGS BRANA STANDARD; PRT; 520 AA.
AC 09M462;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate adenyl1transferase (small subunit, chloroplast
DE precursor (EC 2.7.7.27) (ADP-glucose synthase) (ADP-glucose
DE pyrophosphorylase) (AGPase B) (Alpha-D-glucose-1-phosphate adeny1
DE transferase).
GN AGP51.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_Taxid=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Dreikar; TISSUE=Seed;
RA Zawodny S., Martini N.;
RT "Isolation and analysis of a cDNA clone encoding the small subunit of
RT ADP-glucose pyrophosphorylase in the plastids of seeds and leaves of
RT Oilseed rape (Brassica napus).";
RL Submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN PLAYS A ROLE IN SYNTHESIS OF STARCH.
CC IT CATALYSES THE SYNTHESIS OF THE ACTIVATED GLYCOSYL DONOR,
CC ADP-GLUCOSE FROM GLC-1-P AND ATP
CC -!- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -!- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
CC BY ORTHOPHOSPHATE. ALLOSTERIC REGULATION.
CC -!- PATHWAY: STARCH BIOSYNTHESIS.
CC -!- SUBUNIT: HETEROTETRAMER.
CC -!- SUBCELLULAR LOCATION: Chloroplast (By similarity).
CC -!- TISSUE SPECIFICITY: LEAVES.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-


```

OX NCBI_TaxID=33906;
RN [1]
RP SEQUENCE FROM N.A. STRAIN=CV. FRIBO; TISSUE=Cotyledon;
RX MEDLINE=95218601; PubMed=7766042;
RA Weber H., Helm U., Borisjuk L., Wobus U.;
RT "Cell-type specific, coordinate expression of two ADP-glucose
pyrophosphorylase genes in relation to starch biosynthesis during
seed development of Vicia faba L.";
RL Planta 195;352-361(1995).
CC -I- FUNCTION: THIS PROTEIN PLAYS A ROLE IN SYNTHESIS OF STARCH.
IT CATALYSES THE SYNTHESIS OF THE ACTIVATED GLYCOSYL DONOR,
ADP-D-GULOSE FROM GIC-1-P AND ATP
CC CC
CC -I- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
diphosphate + ADP-glucose.
CC -I- ENZYME REGULATION: ACTIVATED BY 3'PHOSPHOGLYCERATE, INHIBITED
BY ORTHOPHOSPHATE. ALLOSTERIC REGULATION.
CC -I- PATHWAY: STARCH BIOSYNTHESIS.
CC -I- SUBUNIT: HEMERETRAMER..
CC -I- SUBCELLULAR LOCATION: Chloroplast (by similarity)..
CC -I- TISSUE SPECIFICITY: LEAVES AND SEEDS.
CC -I- DEVELOPMENTAL STAGE: IT IS PRESENT IN YOUNG COTYLEDONS AT 14 DAYS
AFTER FERTILIZATION (DAF) WHEN CELLS ARE STILL RAPIDLY DIVIDING.
LEVELS STEADILY ACCUMULATE UNTIL THE END OF THE CELL EXPANSION
PHASE (35-40 DAF) AND WITH THE BEGINNING OF THE SEEDS DESICCATION
PHASE AT 50 DAF, THE LEVELS DECREASE TO VERY LOW LEVELS.
CC -I- SIMILIARITY: BELONGS TO THE BACTERIAL AND PLANTS' GLUCOSE-1-
PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between
the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb.sib.ch).
CC DR EMBL: X76941; CAAS4260.1; -.
DR InterPro: IPR001825; NTP_transferase.
DR Pfam: PF00483; NTP_transferase.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSP_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSP_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase;
KM Multienzyme family; Starch biosynthesis; Allosteric enzyme;
KW Chloroplast; Translt peptide.
FT TRANSIT 1 63 CHLOROPLAST (POTENTIAL).
FT FT 512 GLUCOSE-1-PHOSPHATE ADENYLYLTRANSFERASE
FT SMALL SUBUNIT 2.
FF FE
FO SEQUENCE 512 AA; 56059 MW; 060647f683526EDC CRC64;

```

Query Match 3.2%, Score 101; DB 1; Length 512;
Best Local Similarity 21.0%; Pred. No. 5;
Matches 118; Conservative 68; Mismatches 170; Indels 206; Gaps 30;

Oy	22	P L S S S Q S S D L S Y C S S L P M A S R V T K L V N S A L M T P P A L H F-----PKQSSNSPAIVKP	76
Dh	10	P S S S S S S S S S S S-----SKAIARNLSFTSS-HLSGDKIFLTSGRTTRTSGRNPFVISP	61
Oy	77	K A-KESNTKKOMNLFORAAALDAAAEGLV---SHEKIHLPL-PKTADPSVOIAGFAAYN	131
Dh	62	K A V S D S K S Q T C L O P D A S R V L----GIILGGACATRIYRPLTKRKAKAPVPLGNAVTRL-	116
Oy	132	E Q P V R N L P V V G K L P D S I K G Y V R N G A P L H E P V Y G H F-----PDGDGVNAHV	181
Dh	117	-DIPVSNCNLNSNIISKIYVLTOFN---SASLNRLHSLRAYASNLCGYKNKGFEVEYLA	167
Oy	182	F E-----HGSA-----SYACRETYQN-----RFVQEROLGRPVFPKAIG	215
Dh	168	A Q S P E N P N M F Q G T A D V A R O Y L M L F E E H N V L E Y L I A C D H L Y R M D Y E R-----PIQAHR	221
Oy	216	E L H G H T G I A R L M L F A R A A G I V D P A H G T G V A A G L Y T F N--GRLLANSEDDLPYOVOIT	273

Dd	222	ESDADITVVALPDEEARARA-----		:	:	:	:	259									
Oy	274	PNGD-LK-----TVRFPFDGQLESTMIA-----						301									
Dd	260	PGGEOLAKAKVDTTILIGLDDDAKEKPEPTIASMGIYVSKHWMDLDRKFPGANFGSEV						319									
Oy	302	-PEGSEL-----FALSYD-----						335									
Dd	320	IGATGFELGRVOAYLYDGWEDIGTFEAFYNANMLGITKKVPDEFSEFYDRSSPIYTQP----						376									
Oy	336	EIQLOQPTMMHDPAATENV-----VYPDOQV---VEKLPEMT-----						370									
Dd	377	--RYLPSPKMLADITYDSVIGECCVIKNCKIHHSVVGLSCISSEGAIIDETLLMGADYYE	:	:	:	:	:	434									
Oy	371	-----RCGSPVYVKNNKYARFGILLDKYAEDSSNKKWJADAPCCFHLMANWEDE						420									
Dd	435	TMDRRFLAARKGVPLGICRKNSHIRAITIDKANRIJDVVKIINS-----NWQEARETE						489									
Oy	421	----TDEVVVGSCMTPPDSI 437	:	:	:	:	:										
Dd	490	GXFIKSGIVTVIKDALIPSGTV 511	:	:	:	:	:										
RESULT	13																
ID	ACVS_EMENI	STANDARD:	PRT:	3770	AA.												
AC	P27742:																
Dt	01-AUG-1992	(Rel. 23, Created)															
Dt	01-AUG-1992	(Rel. 23, Last sequence update)															
Dt	16-OCT-2001	(Rel. 40, Last annotation update)															
De	Delta-(L-alpha-aminoacidipyl)-L-cysteiny-L-D-valine synthetase																
De	(EC 6.-.-.-) (ACV synthetase) (ACVs).																
GN	ACVA.																
OC	Emeritella nidulans (Aspergillus nidulans).																
OS	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;																
OC	Eurotiiales; Trichocomaceae; Emeritella.																
RX	NCB1_TaxID=5072;																
RA	[1]																
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.																
RC	STRAIN-G191;																
RX	MEDLINE=91286299; PubMed=2061333;																
RA	Maccabe A.P., van Liemp H., Fallissa H., Unkles S.E., Riach M.B.R.,																
RT	Pfeifer E., von Doehren H., Klinghorn J.R.;																
RT	"Delta-(L-alpha-aminoacidipyl)-L-cysteinyl-D-valine synthetase from																
RT	Aspergillus nidulans. Molecular characterization of the acva gene																
RT	encoding the first enzyme of the penicillin biosynthetic pathway.";																
RL	J. Biol. Chem. 265:12646-12654(1991).																
CC	-I- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF THE TRIPEPTIDE																
CC	ACY ARE ACTIVATED AS AMINOACYL-ADENYLATES WITH PEPTIDE BONDS																
CC	FORMED THROUGH THE PARTICIPATION OF AMINO ACID THIOLESTER																
CC	INTERMEDIATES.																
CC	-I- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHINES																
CC	(POTENTIAL)..																
CC	-I- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND																
CC	CEPHALOSPORIN.																
CC	-I- PTM: THE N-TERMINUS IS BLOCKED.																
CC	-I- SIMILARITY: BELONGS TO THE AMP-DEPENDENT AMP-BINDING ENZYME																
CC	FAMILY.																
CC	-I- SIMILARITY: CONTAINS 3 ACTL CARRIER DOMAINS.																

DR InterPro: IPR001242; DUF4.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR003880; Phosphopant_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 3.
DR Pfam: PF00668; Condensation; 3.
DR Pfam: PF00550; pp-binding; 3.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 3.
DR PROSITE: PS00455; AMP_BINDING; 3.
DR PROSITE: PS50075; ACP_DOMAIN; 3.
DR Liasse: Antibiotic biosynthesis; Multifunctional enzyme;
Repeat; Phosphopantetheine.
FT REPEAT 321 910 DOMAIN 1 (ADIPATE-ACTIVATING).
FT REPEAT 1413 1993 DOMAIN 2 (CYSTEINE-ACTIVATING).
FT REPEAT 2494 3078 DOMAIN 3 (VALINE-ACTIVATING).
FT DOMAIN 850 919 ACYL CARRIER (ACP) 1.
FT DOMAIN 1929 2002 ACYL CARRIER (ACP) 2.
FT DOMAIN 3020 3087 ACYL CARRIER (ACP) 3.
FT BINDING 882 882 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 1965 1965 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT BINDING 3050 3050 PHOSPHOPANTETHEINE (BY SIMILARITY).
FT ACQ_SITE 3623 3623 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 3770 AA; 422448 MW; CB66BD232A58CB0 CRC64;

Query Match 3.2%; Score 100; DB 1; Length 3770;
Best Local Similarity 20.9%; Pred. No. 1.le+02;

Matches 97; Conservative 67; Mismatches 162; Indels 138; Gaps 27;

QY 12 GMLGNTPTPLSSSSDLSYCSLPMASRYTRK-----LNVSAALTPPAL 60
DB 1544 GMSWSN-----PAPSTSTDLAY-----IITYSGTGPKGVMEHHGVNWLQSLSTFEL 1596
QY 61 -----HPPKSS-----NSPAIVVKPKAKESNTKOMLEPRAAALDA 99
DB 1597 RDTDEVILSFSNYVDFHVEQMTDALINGQTLVMLNDAMRSDKER--LYQYLETNRVTV 1654
QY 100 AEGFLVSEKHLRLPKTADPSVQIAGNFAPVNEQPVRR-----NLVYVKRLDPSISG 151
DB 1655 LSG-----TPSYISMVFEFSRFRKH--LRRVDVCGEAFSQVPEQDIRTFQG 1698
QY 152 VYVRGANGPLHEPVYTGHHFFDGDGMVHAKVFEHGSASYACRFQTNRFQVORQLGR-PVF 210
DB 1699 LII-NGYGTETISITTHK-----RLVPRPERRTDKSIGQIGNSTSYLNDMKNRPV 1750
QY 211 PRAIGELH-GHTGIRALMLFTAR-----AAAGIVDPAHGTGVANAGLVFNGRLANSDD 265
DB 1751 -GAVGELYLGEGVAR--GYNHRPEVTAEERFLRNPFOJIDSEKONGR--NSRLYRTGD-- 1802
QY 266 LFOYOIIP--NGDKTYGRFDFD-----GOLSTMTAHP-----KYDPES 304
DB 1803 ---LYRWIPGSGEIEYIGRNDFOVKIRGLRIELGIEIAVMSHPDIKOSVVIANSKGEG 1859
QY 305 GELFALSTDVYVSKPYLKYFRSPDGTSPDVEIQIDQPTMMDFAITENFV-----VVP 358
DB 1860 DQKFLVGVFAVSSP-----LSPGA-----IRRMQSRRLGVMPPSFPISSLPVPR 1906
QY 359 DQGVVFK-LPEM-IRGGSFVYVDKKNVAF-----GIIDKTAE 394
DB 1907 SGKLDITKALPTAEKGMANVLAPRNEIESILGISAGLIDISAO 1950

RESULT 14

PHYC_BACSU STANDARD; PRT; 383 AA.

AC 031097;
DT 30-May-2000 (Rel. 39, Created)
DT 30-May-2000 (Rel. 39, Last sequence update)
DE 30-May-2000 (Rel. 39, Last annotation update)
DE 3-phytase precursor (EC 3.1.3.8) (Phytate 3-phosphatase) (MIO-inositol-hexaphosphate 3-phosphohydrolase).

GN PHYC OR PHYB13.
OS Bacillus subtilis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-VTT E-68013;
RA MEDLINE-98268943; PubMed-9603817;
RX Kervuo J., Laureus M., Nurminen P., Kalkkinen N., Aapalahti J.,
RT "Isolation, characterization, molecular gene cloning, and sequencing
of a novel phytase from Bacillus subtilis."
RT Appl. Environ. Microbiol. 64:2079-2085(1998).
CC - FUNCTION: CATALYZES THE HYDROLYSIS OF INORGANIC ORTHOPHOSPHATE
FROM PHYTATE. ONLY PHYTATE, ADP, AND ATP WERE HYDROLYZED (100, 75,
AND 50% OF THE RELATIVE ACTIVITY, RESPECTIVELY). HAS MAXIMAL
ACTIVITY AT PH 7 AND 55 DEGREES CELSIUS.
CC - CATALYTIC ACTIVITY: Myo-inositol hexakisphosphate + H(2)O = D-myo-
inositol 1,2,4,5,6-pentakisphosphate + phosphate.
CC - COPACITOR: ISOLATED ENZYME REQUIRED CALCIUM FOR ITS ACTIVITY AND/OR
STABILITY.
CC - SUBCELLULAR LOCATION: Secreted.
CC - INDUCTION: BY PHYTATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF029053; AAC31775.1; -.
DR HSRP: O66037; 2PRO.
DR InterPro: IPR003431; Phytase.
DR Pfam: PF02333; Phytase; 1.
KW Hydrolase; signal.
FT SIGNAL 1 26 POTENTIAL.
FT PROPEP 27 29
FT CHAIN 30 383 3-PHYTASE.
SQ SEQUENCE 383 AA; 41922 MW; E9BEC24A48B99CA CRC64;

Query Match 3.2%; Score 99.5; DB 1; Length 383;
Best Local Similarity 19.3%; Pred. No. 4.2;

Matches 88; Conservative 71; Mismatches 151; Indels 147; Gaps 22;

QY 18 NHTQPLSSSSQSDLSYCSLPMASRYTRKLN-----VSSALHTPPALHPPKSSNSPA 71
DB 2 NNSKTLLTLTAAGIMLTGCA--VSSQAKHKLSDPYHFTVNAAEETEP-VDTAGDAADDP 58
QY 72 IYVVKPKAKESNTKOMLEPRAAALDAAGFLVSEKHLRLPKTADPSVQIAGNFAPV 131
DB 59 IWDPRTPQ-NSR-----LITNKKSGLVVYSLDGMKLSYNTGKLN 99
QY 132 EQVPRNPLVYVKLPDPSIKGVYVRNGANGPLHEPVYTGHHFFDGDGMVHAKVFEHGSASYAC 191
DB 100 NVDIRYDFPLNKKVDIAAASNRSECKNTIE-----IYALDGKNGTL----- 141
QY 192 RFTQTNRFQVORQLGRVPR--KAIGELHGTGIRALMLFYARAAGIVDPAHGTGVANA 249
DB 142 -----OSMEDPRDPIATAIINEVGYFT-----LYHSQK----- 168
QY 250 GLVYFNGRLAM--SEDL-FOYOITPNSGL--KTGRGFRPDGOLSTMTAHPKVDE 303
DB 169 -----TGKTYAANTGKEGFEBOIELADKNGYISGKVAFAKNSOTEGM-----AADE 218
QY 304 SEELFALSTDVYVSKPYLKYFRSPDGTSPDVEIQIDQPTMMDFAITENFV-----ATTENFV 355
DB 219 YGRLLYAEDEA-----IMKFSAPREGSGNGYIDRADGNHLIRDIGLITTYAAADGKYL 274
QY 356 VVPDQGVVFKLEPMIRGGSFVYVD--KKK-VAREGILDKYEDSSNIMWIDAPDCFCFH 411

DB 275 MASSQ-----GNSSYAIVDRQKNKYVDFRITDGPETDGT----- 311
QY 412 LMANEPEPTEDEVVIGSCMP--PDSEFNESE-ENL 445
DB 312 -----DIDGIDVLGRGLGPEYPRGIFVADGENI 340

RESULT 15
ATL_STRAU ID ATL_STRAU STANDARD; PRT: 1256 AA.
AC P52081;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Bifunctional autolysin precursor (includes: N-acetylmuramoyl-L-alanine
amidease (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-
acetylglucosaminidase (EC 3.2.1.1.96)).
GN ATL.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC STRAIN-RN450;
RX MEDLINE=95116542; PubMed=7816834;
RA Oshida T., Sugai M., Komatsuzawa H., Hong Y.-M., Suglnaka H.,
RA Tomasz A.;
RT "A. Staphylococcus aureus autolysin that has an N-acetylmuramoyl-L-
alanine amidease domain and an endo-beta-N-acetylglucosaminidase
domain: cloning, sequence analysis, and characterization.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-MCTC 8325-4;
RA Foster S.J.;
RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLGHIOTOBIOSYL UNIT IN
CC HIGH-MANNOSE GLYCOPOLYMERIDES AND GLYCOPOLYMERIDES CONTAINING THE
CC -[(MAN)5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYLD-D-GLUCOSAMINE
CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC OLIGOSACCHARIDE IS RELEASED INTACT.
CC -1- CATALYTIC ACTIVITY: Hydrolyses the link between N-acetylmuramoyl
CC residues and L-amino acid residues in certain bacterial cell-wall
CC glycopeptides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC EXTRACELLULAR LYTIC ENZYMES.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D17366; BAA04185.1; -
DR EMBL; LA1499; AAA99982.1; -
DR InterPro: IPR002502; Amidase_2.
DR InterPro: IPR002901; Amidase_4.
DR Pfam: PF01510; Amidase_2; 1.
DR Pfam: PF01832; Amidase_4; 1.
DR SMART; SM00047; LY22; 1.
KW Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
FT SIGNAL 1 29
FT CHAIN 30 1256 POTENTIAL.
FT DOMAIN 199 775 BIFUNCTIONAL AUTOLYSIN.
FT DOMAIN 776 1256 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT DOMAIN ENDO-BETA-N-ACETYLGUCOSAMIDASE.

FT REPEAT 425 589 1.
FT REPEAT 596 758 2.
FT REPEAT 770 932 3.
SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;

Query Match 3.2%; Score 99.5; DB 1; Length 1256;
Best Local Similarity 19.3%; Pred. No. 24;
Matches 89; Conservative 59; Mismatches 164; Indels 149; Gaps 20;

QY 7 TAAYSGRWLGG-----NHQOP-----PLSSOSSDLSYC-----SLPMASRY 44
DB 98 TAQVNDGDTFRANQATNTNNQPVAKSTSTPAKTNTVTNAGSLVDEDEDNSNQINPEL 157
QY 45 FRKLNVSSALHTPPALHPKROSSNSPAIVYKPAKESNFKOMLFGRAAAALDAAEGL 104
DB 158 IKSAAKPAALETQYKTAAPKAATTS-----APRAKTEALPKYTFESASAPRSVAA----- 208
QY 105 VSHEKLHPLEKTPADPSVOIAGNFAVNEQVRRNLEPV-----VGKL 145
DB 209 -----TPKTSLPKYPKQVN--SSINDYICKNNLAKAPKIEDYTSYPPKYAYRNGVGR- 258
QY 146 PDSIKGYVVRANGANPLHEPYTG-----HHFFDGGWVHAAYKFEHGSASYAC 191
DB 259 PE---GIYVHDTAND--RSTINGEISYMKNNYQNAFVHAFAVDGRIIETAPTDYLSGVGA 314
QY 192 RFTQTRRFVQEROLGRPVPPKA--IGELHGTGIGARLMLFYARAAGIVDPAHGTGVANAG 250
DB 315 -----VGNPRFINVEIVHTHDYASAPRSNNYADYAA-----TQ 348
QY 251 LVYFNGRLAMSBDLLPYQVITPNGDLKT--VGRF-----DFDQLESTMIAPRY 300
DB 349 LQYYGLK-----PDSAEYDNGTVMTHYAVSKYLIGCTGDHADPHGYLRSHNSYDOL 399
QY 301 DPESGELFALSYDVVSKRYL-KYFRSPDGTSPDVIEQLDPTMHHDRATNFVVPD 359
DB 400 -----YDLINELKYLIMKGVAPMCTOSTPTPTTPSKPT-----TPS 435
QY 360 QQVVPKLPKEMIRGSPVVDKKNVAFRGILDKYAEDESSNIK 400
DB 436 KPSTGKLTVAANNQVQAIKPTNSGLTYTYDRIKGAATNEQ 476

Search completed: July 19, 2002, 10:16:06
Job time: 233 sec

Db 277 DCFESNETMGWLHIADKKRKYLNKRYTSSFNT-----FHHINTYEDNEF--L 324
Qy 425 VVIGSCMPDPDSIFNESP-ENLKSYLEIRLNKCTGSTRPRLIISNDOQVLEGMVR 483
Db 325 IYDLCQWKGFEFVYNYLYLANLRNMEEVK---KNARKAPQPEVRSVPLNIDKADYK 381
Qy 484 NMLGRKTKFAYIAL-----AEPMPKYSG-----FAKVDLTT--GEVKKHLYG----- 523
Db 382 NLVTLPTATATATLSDETIMLEPEVLFSGPRQAFEPQINQYGGKPYTYAAGLGNH 441
Qy 524 -----DNRYGGEPLFLP--GEGGEDEGYILCFVHDEKTK--SE 559
Db 442 FVPDRCLKLVNKTETWQOEPDSIPSEPIFVSHPDALBEDDGVLSVVSFGACOKPAY 501
Qy 560 LOIVNAVSLVEEATVKLPSPVRYFGHGT 588
Db 502 LLILNAKDLSEVARAEVEINIPVTFHGLF 530

RESULT 2
US-08-488-305A-6
; Sequence 6, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B VIK, Claes Olof, Eriksson, ULF, Peterson, Per A.
; TITLE OF INVENTION: Isolated protein Receptors, Antibodies Which
; TITLE OF INVENTION: blind thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Theretofor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488, 305A
; FILING DATE: 7-JUNE-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Kohli, Vinset
; REGISTRATION NUMBER: 37,003
; REFERENCE/DOCKET NUMBER: LUD 5280.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-488-305A-6

Query Match 6.6%; Score 209; DB 1; Length 533;
Best Local Similarity 20.7%; Pred. No. 6.1e-13;
Matches 119; Conservative 99; Mismatches 222; Indels 134; Gaps 29;

Qy 105 VSEKLIHPLRKADPSVQIAGNFAPVNRPNLP-----VVGKLPDSIKYVYVNGANP 160
Db 1 MSSQVEHP-----AGGYKLFETVEELSSPLAHVGRPLMLWTGSLLCGPGL 49
Qy 161 LH---EPVYGHFFDGDGVNVAHKFEHGSASYACRFTQTRFVQ-----ROLGRPV 209
Db 50 FEVGSSEPF--YHLPDGOALLHKFKEGHVYTHRRFIRIDAVYRAMTEKRIYITEFGICA 107

Qy 210 FPKAIGELHGHGTIARLMLFYARAAAGIVPAHGTGVANAGLVYENGRLLAMSED----- 264
Db 108 FPDPCKNI-----FSRFFSYF-----RGVEVTNALV-----NLYPGEDYACT 147
Qy 265 DLPYQVQITPNQDLTKVGFDPDGOLE-STMAHKKVNPES-----GELFALSTDV 314
Db 148 ETNFTTKVNPETLTETIKQVDCNYSVNGATAHPIENDGTVYVNIQNGCFGNFSIAYNI 206
Qy 315 VSKPYLYKFRFSPDGRKSPDVEIOLD-----QPTMHDAITENPVYVPPDQOVFKLPM 369
Db 207 VAIPLQADKEDP--ISKSEIYVQPCSDRFRKPSYVHSGLPNTIYVETVYKINLEKF 264
Qy 370 IKGSPVYVYDKKVAARF-----GLDKRYAEDSSNMIKIDAPDCECFILMANAEP 419
Db 265 LSWMS--LWGANVMDCEESNETMGWLHIADKKRKYLNKRYTSP-FULFHHINTYEDH 321
Qy 420 ETDEVVYIGSCMTPDPSIENSD-ENLKSYLEIRLNKCTGSTRPRLIISNDOQVNLFA 478
Db 322 EF--LIVDLCQWKGFEFVYNYLYLANLRNMEEVK---KNARKAPQPEVRSVPLNIDK 376
Qy 479 GMYNRNM--LGRKTFAYIALAEP-W--PKV--SG-----FAKVDLTT--GEVKKHLYG 523
Db 377 ADTGKNLVLPNTATATLCSDETIMLEPEVLFSGPRQAFEPQINQYGGKPYTYAYG 436
Qy 524 -----DNRYGGEPLFLP--GEGGEDEGYILCFVHDEKTK 556
Db 437 LGILNFVDPRLCLKLVNKTETWQOEPDSIPSEPIFVSHPDALBEDDGVLSVVSFGAG 496
Qy 557 K--SELQIYNAVSLVEEATVKLPSPVRYFGHGT 588
Db 497 QRPAYLLIIMADLSEVARAEVEINIPVTFHGLF 530

RESULT 3
US-08-928-361B-30
; Sequence 30, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, CARYN, JONES & BIKSA
; STREET: 365 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928, 361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026, 062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: VERNY, HANA
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1043 amino acids

```

;      TYPE: amino acid
;      STRANDEDNESS:
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-928-361B-30

```

Query Match	3.9%	Score 123.5	DB 3	Length 1043
Best Local Similarity	20.6%	Pred. NO. 0.0018		
Matches 125, Conservative	85	Mismatches 194	Indels 203	Gaps 35

[illegible]

```

RESULT 4
US-08-700-651-5
Sequence 5, Application US/08700651B
Patent No. 6015882
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JIRI
TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
FILE REFERENCE: 480.19-4(HV)
CURRENT APPLICATION NUMBER: US/08/700,651B
CURRENT FILING DATE: 1997-08-14
EARLIER APPLICATION NUMBER: 08/415,751
EARLIER FILING DATE: 1995-04-03

```

```

; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1721
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
US-08-700-651-5

```

Query Match	3.9%	Score 123.5	DB 3	Length 1721
Best Local Similarity	20.6%	Pred No. 0.0044		
Matches 125, Conservative	85	Mismatches 194	Indels 203	Gaps 35

```

0Y 5 TATAAASRWMLGNHTOPPLSSOSSLSYSCSS--PMARVYRK----LWSSALHTPP 58
   ||| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 831 TIAGIVSG-----ISASESL-LSQKSALIDPATMMVYGEFGGLLNPATGWIIRG 878
0Y 59 ALHFPKQSSNSPAT---VVKPKAKESNTKOMNL-----FORAAAAAD 98
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 879 FLGSEQOFOFSPLEIDEGGILPEVVAANAKFKLISPPESPIPEKDKIDISISELMTY 938
0Y 99 AAEGLVSEHELHLPLKPTADPSVQIAGNFAPVNGQPRRLPVYGGKLPDSIKKVVYRNCA 158
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 939 IESGGLIGOVSKRRLPGS-----IADGLNIMTPTOTD-SYTGKPIDTTL 985
0Y 159 NPLHEPVYGHFFPDGDGVNAHVKKFEHGSASY--ACRFTQTNREVOEKOLRPFYPAKIGE 216
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 986 -PENPR-TGH-----LINPTNNTMDSSPAGAYKKAASNGIKTDVYGLPV-----GE 1031
0Y 217 LHGTGTIARLMLFYARAAGIYDPAHGTVAN--AGLVYNGRL-LAMSEDDLPLYOVQIT 273
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1032 ITGLPKDQSDIPENSTTGLVDESTGKPIINSTAGIV--SGKRGPLRPIDE----- 1081
0Y 274 PNGDLKTYGRDP-----FDGQLESTMIANHKVYDSESGELFALSYDVVSKPPLYKYREFSP 327
   ||| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1082 -NGNL-----FDPSTNLPIDGN-----NOLVPEPNTSVSGSTGTTKP-----KP 1121
0Y 328 -----DGTKSDVEI--QLOD-----PTMMHDPATENFVVVPDQOYVE 364
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1122 GIPVGGGVVPDEEAKDQADKGDGLVPTPTSINKPVTNYQTSGNTGMIINPETGKV- 1180
0Y 365 KLPENIRG-----GSVVYVDKMKVAFGLDKVYAADSSNIKMIADPFCFLHNAME 417
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1181 -IPGSLPESLNVPSFNFQOTDE-----ITGKPRVTVYGLPV----- 1216
0Y 418 EPETDEVVIVIGSCMTPPSIFNESDENLKSYSLEIRLNLKTGESTRRPIISNEDOQVLE 477
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1217 DPTGCEIITDPAKTLPIRPSVAGD-----ELTREV-LNITTTDEVYGLPI-----DLE 1261
0Y 478 AGVAVNRMLGKRTKFAVYALALAEPPKVVSGEAKV--DLTTGEYVKHLVGDNRVYGGEBPLT 544
   :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1262 TGL-----PRDPVSGLPQLPNCGLTVDPNSKPKIPGS--HSG--FI 1297
0Y 535 PGEEGEE 541
   ||| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 1298 NGTSGEQ 1304

```

```

: RESULT 5
: US-08-928-361B-6
: Sequence 6, Application US/08928361B
: Patent No. 6071518
:
: GENERAL INFORMATION:
: APPLICANT: Petersen, Carolyn
: TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
: TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
: TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
: TITLE OF INVENTION: SPECIES INFECTIONS
: NUMBER OF SEQUENCES: 30
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PETERS, VERNY, JONES & BIKSA
: STREET: 385 Sherman Avenue, Suite 6
: CITY: Palo Alto
:

```


STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Vermy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEO ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1721 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-928-361B-6

Query Match 3.9%; Score 123.5; DB 3; Length 1721;
Best Local Similarity 20.6%; Pred. No. 0.0044;
Matches 125; Conservative 85; Mismatches 194; Indels 203; Gaps 35;

QY 5 TPAVAGSMWLGNNHTQPLSSQSSDLSYCSL--PMASTVTRK---LWSSALHTPP 58
DB 831 TIAGIVSG-----ISASESL-LSQKSLIDPATNMVYGGGLNPATGVIMIG 878
QY 59 ALHFPPKSSNSPAI-----VVKPKAKESNTKQNL-----FQRAAAALD 98
DB 879 FLGPSEQOTPFSPEDIDGIIIPPEVAANADKFKLSIPSPVPESIPKDKIDISIELMD 938
QY 99 AAEGLVSHKHLPLPKTADPSVOIAGNFAVNEQVRRNLPPVYKLPDSIKGVYRNQA 158
DB 939 IESGRILQGVSKRPIGS-----IAGDLNPKPTQTD-SYTGKRPIDPTTGL----- 985
QY 159 NPLHEPVYTGHHFFPDGDGVAHVAFKEHGSASY--ACRFOTQNRFOERQLGRPVFKAIQE 216
DB 986 -PFNPP-TGH-----LINPTNNNTMDSFAGAYKYAVSNGIKTDNVYGLPV-----GE 1031
QY 217 LHGHTGIARLMLFYAANAAGIYDPAHGTGVAN--AGLYVFNGRL-LAMEDDLPIYOVT 273
DB 1032 ITGLPKPDPSDIPFNSITGELVDPSTGKPIINNSTAGIV--SGKPGGLPIEDE----- 1081
QY 274 PNGDLKTVGRFD-----FDGQLESTIMIAHPKVDESGELFALSIVYVASKPYLKYFRFSP 327
DB 1082 -NGNL-----FPPSTNLPIDGN-----NQLVNPETNSTVGSSTGCTKP-----KP 1121
QY 328 -----DGTSPDYET--QLDQ-----PMMHDFAITENFVVPDQOQVF 364
DB 1122 GIPIVNGGVVPPDEAKDQADKGDGLIYPPTNSINKDPYNTNOYSTNTGNIINPERGKV- 1180
QY 365 KLEPMTIRG-----GSPVYVDKNKVARFGILDKYAEDSSNKKMIDAPCCFCHLMNAME 417
DB 1181 -IPGSLPGSLNTPSENFPOOTE-----ITGKPVIVYTGILPY----- 1216
QY 418 EPEETDEVVYIGSCMPDPSIFNESDENLKSVLSEIRLNKLTGSESTRPIISNEDQOVNLE 477
DB 1217 DPTGTETIDPATKLPPIPGSVAD-----EILLEV-LNTITDEVYICLP-----DLE 1261
QY 478 AGVNNRMILGRKTKFAYLALAEPMKVSFAKY--DLTTGEVKKHLYGDNRYGGEPLFL 534

DB 1262 TGL-----PRDPVSGLPOLPNCTLVDPNSKKRPIGS--HSG---FI 1297
QY 535 PGEGEE 541
DB 1298 NCTSGEQ 1304

RESULT 6

US-08-928-361B-11
Sequence 11, Application US/08928361B
Patent No. 6071518

GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIRSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:

NAME: Vermy, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480,76-1(HV)
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEO ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 1042 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-928-361B-11

Query Match 3.7%; Score 116.5; DB 3; Length 1042;
Best Local Similarity 19.9%; Pred. No. 0.0098;
Matches 113; Conservative 74; Mismatches 175; Indels 207; Gaps 30;

QY 48 LNVSSALHTPPPLHPKQSSNSPAI-----VVKPKAKESNTKQNL----- 88
DB 190 LNPATGVMLPGSLGPSEQTFPSPEDIDGIIIPPEVAANADKFKLSIPSPVPESIPKDK 249
QY 89 -FQRAAAALDAAGFLVSHKHLPLPKTADPSVOIAGNFAVNEQVRRNLPPVYKLPD 147
DB 250 KIDSISELMYDIESGRILQGVSKRPIGS-----IAGDLNPKPTQTD-SYTGKRPID 302
QY 148 SIKGVYVRNGANPLHEPVYTGHHFFPDGDGVAHVAFKEHGSASY--ACRFOTQNRFOERQL 205
DB 303 PTTGL-----PFNPP-TGH-----LINPTNNNTMDSFAGAYKYAVSNGIKTDNVY 347
QY 206 GRPV-----FPAKIGELGHGTGIARLMLFYAANAAGIYDPAHGTGVAN-- 248


```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2172 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-611-107-31

Query Match      3.3%, Score 104.5; DB 1; Length 2172;
Best Local Similarity 19.9%; Pred. No. 0.65;
Matches 118; Conservative 80; Mismatches 247; Indels 147; Gaps 27;

OY 66 SENSIAIVKPKAKSNKOMLFCORAAALDAEEGLYSVHEKLHPLPKTADSVOIAG 125
Db 1406 TGNCTGVDIRVEDPNTHK--LFRSATPTGAPLHG-IALHEPYKL-----DAIDLR 1457
OY 126 NEPAVNEOPVRRLNPFVGKL-----PDSIKVVYRNGANPLHEPVTCGHFFDDGM-- 176
Db 1458 AARARNETTYCDPEFLAETALAKSKWESGISHVASESNEHORAYAEVELIFADSTGSWG 1517
OY 177 -VAHKFEHSGASVCRTQTNRNFQEQSLRPVF-----PRAIELGHT 221
Db 1518 PLVPERPPGSINCEVAMNNKLSPPERGGREILVYANDYTFAGSGCPREDAFPPAVT 1577
OY 222 GIA---RLMIFYARAAGIYDPAHGTGVANAGLYFENGRLAMSDDLPR---XVOYTP 274
Db 1578 NLACERKIPLYLSTAGA-----RLGAEEIKACFH---VGWEDDQSPEGFYIYLTE 1629
OY 275 NODLKTVGRFPEDGLESTMTAHPKVDESEGEELALSVDYVK----- 317
Db 1630 Q-----DY-SRLSSSVIALHELKYPESEGSETRWVDTIVKKDKDGCCENLMGSAIAS 1679
OY 318 PLYKYFR--FSPDGKSPDVEI-----OLDOPMHMDFAITENFVVVPDOOV 362
Db 1680 AYSAKYREFLTFTYTGRAIGIAGVALRLMGRCIORLDQPIILGYSAKNKL---GREV 1736
OY 363 VFKELEMTIRGSPVYYDKNKYARFGILDKAIEDSSNIKWIDADPCFPHLMAMEEPTD 422
Db 1737 YSSOMQL---GGPKIMATNGVGHLLVSDDLEGSAIIKWL-----SYPPRYVG 1781
OY 423 EYVVVSGCMTPPDSTIFNSDENLKSVLSEIRLNKTGSTRRPFIISNDQVNLEAGMVN 482
Db 1782 GPLPYVSKLPDPERAVTYFFPEN-----SCDARAALICGIODTQGKMLSGMFD 1827
OY 483 RNLMLGRKTKFAYVLAALAEPMK-VSGFAKYDLTTGEVKKHLVGDNRYCGEPFLPG---- 536
Db 1828 RE-----SFVETLEGMATKITVTIGRKASAATAQA---LLDENR-EELPLFITLAMRG 1875
OY 537 -EGGED--EGYI---LCFVDEKTKWE--LOIVNAVSLVEEATVRLPSRV 580
Db 1876 FSGGORDLEGILOGAKMITVENLTRYKOPAFVIYPKAGELGKAMVVVDSKI 1927

RESULT 10
US-09-001-984C-106
; Sequence 106, Application US/09001984C
; Patent No. 6245331
; GENERAL INFORMATION:
; APPLICANT: Laal, Sumun
; APPLICANT: Zolla-Pazner, Susan
; APPLICANT: Bellisle, John T
; TITLE OF INVENTION: EARLY DETECTION OF MYCOBACTERIAL DISEASE
; FILE REFERENCE: NYU-011
; CURRENT APPLICATION NUMBER: US/09/001,984C
; PRIOR FILING DATE: 1997-12-31
; PRIOR APPLICATION NUMBER: 60/034,003
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 741
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
```

US-09-001-984C-106

Query Match 3.1%; Score 98; DB 4; Length 741;
Best Local Similarity 21.7%; Pred. No. 0.47;
Matches 139; Conservative 79; Mismatches 216; Gaps 37;

QY	97	LDAAEGFVLSHEKHLPLP-----KTADPSVOAGNFAPVNDOPV---RRLPVV-----GK	144
Db	78	MDAIRQFPLETELGILLPEPDDPTTITTSQVDAEITTTAGPQLVWVPLNARFLMANNAWGS	1377
QY	145	LPDSIKGVYV---RNGA--NPLHEPVYV-----HHFFPDGDGVNAHVAKKEHGSASYACR	1922
Db	138	LYDALYGGDVIPETDGAKEGPTYKKVYBGDKVIAVARKFLD-----DSVPLSSGFGATG	1922
QY	193	FTQJNRFVQEOQLORPVYPPKAI-----GELHGTGILR-----LMFYARAAGIYDPAH	2422
Db	193	FT-----VQDQLVVALPDKSTGTLANPGOFAGYGAASEPTSVLLIHHGLHIELIDPES	2474
QY	243	GTGVAN--AG-----LVYFNGRLLAMSEDD--LRYOYITPN--DDLKTIV-----	281
Db	248	QVGTITDRAGVADVILLESAITTIMPEEDSVAAVDAADVYLGRMKGLINKGDLAANDXOG	3072
QY	282	-----GREDFDQO-----LESTMIHAPKVDPESGELFALSYDVS	316
Db	308	TAFLRLVLRDENTAPYAGGGCTTLRGRSLMEFRVNGHILMTDAIVDTDGESEFEGIMDLF	367
QY	317	KPYLKYFFPSDGTGKSPV-----ELQDQPTMMH---DFALT--ENFVVPVPPQOY	362
Db	368	TGLI-----AHHGLKASDVNGPLLSRTGSITYYKPK--KMGPAVAVATCELFSEVED--	418
QY	363	VFKLPEMIRGGSPVYVYDNKNKVARREGLIDKYAEDSSNJKWIDAPDCFCFHLMNAMBEPTD	4222
Db	419	VLGEPQ-----NMKIGIMDEBERTVNLK-----ACIKAAD	451
QY	423	EYVVVIGSCMPDPDSIENBSEDNKLSVSEIRLNKJTGSESTRP-----IISNDOQVN	475
Db	452	RYVEINTGF-----LDRTDGEIHTSMEAGPVMYRKGSQPMILAEEDHND	498
QY	476	--LEAGVNRMLGR-----KTRFAV--LALAPW--PKYSG-----PAKY	510
Db	499	AGLAAEGSGRAOVKGMKMTMTELMADVETKIADPRAGASTAWPSPATLALAHNOY	558
QY	511	DLTTGVEVKNHLYGNRRYGEPL-----FLPGEGBEDE-----GYLICEVUDE	553
Db	559	DVAAN--VQGLAGKRRAITBQLLITPLAKELAMAPDEIREVDNOCOSIIIGYVBRW--DQ	615
QY	554	KTKMSELDIYNAVSL--EVEATVKLPSRYPGE--HGFFIGAD	592
Db	616	GVGCSKVDPDIDHVALMEDRALRISOLLAAMWLNHGVTSD	657

```

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1
```

```

1      SOFTWARE: PatentIn Release #1.0, Version #1.25
2
3      CURRENT APPLICATION DATA:
4      APPLICATION NUMBER: US/08/097,829
5      FILING DATE: 19930723
6
7      CLASSIFICATION: 435
8
9      ATTORNEY/AGENT INFORMATION:
10     NAME: Heslin, James M.
11     REGISTRATION NUMBER: 29,541
12     REFERENCE/DOCKET NUMBER: 1216-35
13
14     TELECOMMUNICATION INFORMATION:
15     TELEPHONE: 415-326-2400
16     TELEFAX: 415-326-2422
17
18     INFORMATION FOR SEQ ID NO: 4:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 516 amino acids
21     TYPE: amino acid
22     TOPOLOGY: linear
23
24     MOLECULE TYPE: protein
25
26     OS-08-097-829-4

```

Query Match	3.1%	Score	97	DB-1	Length	516
Best Local Similarity	20.9%	Pred.	No. 0.32			
Matches 123; Conservative			69;	Mismatches	172;	Indels 224; Gaps 32

```

QY 1 MASFTATAVSGRWLGGHHQHPRLSSOSSDLSCSSLPMASRYTRKLTNYS- L 54
Db 1 MASAALICVL- - - - -VPSSSSSS- - - - -SSSSSKALARNLSTSSQLCGDKT 46
QY 55 HTTPALHPKROSSNSPAIVWPKA-KESNTPKQMLFPRAAAALDAIEGVL--SHEKL 110
Db 47 FT--VSGTRRSSGRNRPITVSPKAVSOSKQNCQICLDPDASVYL--GIILGGAGTRL 99
QY 111 HPL-PTADPSVQIAGNEFAPVNEOPVRBNLPVYGLKLPDSIKGVYRNGANPLHEBVTGHH 169
Db 100 YPLTKRAKPAVPLGANRLT- - - - -DLPVSNCLNSINISKIYLTQFN--SASLNRH 149
QY 170 F- - - - -FDODGMYHAKFE- - - - -HGS- - - - -STACRTQTN- - - - - 197
Db 150 LSRAYASNLGGYKKEGFEVYLAQAQSPENPNWFGOTDAVRQYIMLEEEHNVLLEYLVLG 209
QY 198 - - - - -RFVQERQLGRVPVPKAIGELHGTGLARLMLFAPRAAGLVDPRAHGTGVANAGLV 253
Db 210 DHLRYMDYER- - - - -FLQALHRESADADITVASLPMDEARATA- - - - -FLGIM 250
QY 254 FN--GRLLAMSE- - - - -DD- - - - -LFPYQVQITPENG- - - - -D 277
Db 251 IDEGRIIVEFSEKPKGEOLKAMKVDTTLLGDDBERAKEMPR- - - - -IASMGIYVSKHYMD 307
QY 278 L--KTVRGFPDQLEBSTMIAHKVDPESGL- - - - -FALSYD- - - - - 313
Db 308 LLRKDFPGANDFGSB- - - - -VIPGATELGLRVAALVLYDYMEDIGTIEAFYANL 357
QY 314 -VVSXP--YLYKFRFSPDGTKSPDVEIOLDOPMHMDFAITENFV- - - - -VWPDQOV- - - - - 362
Db 358 GIYTKKPVDFESFYDRSSFIYQP- - - - -RYLPFSKMLDADITBVDVJEGCYINCKTHHS 412
QY 363 VFKLPEMI- - - - -RGSGFVYVYDKKKVAFGLDKYAE 394
Db 413 VYGLASCSISEGATIEDTLMGADYLETIDADRFLAAGVPIGIGKNSHKRALIDKRNAR 472
QY 395 DSSNKTWADADCCFHLHMNMEPE- - - - -TDEVVYVIGSCMPPPDSI 437
Db 473 IGDGKIINS- - - - -NVQELARTEGKFIKSGVLYIKALKLPPSGV 515

```

RESULT 12
US-08-577-403-4
; Sequence 4, Application US/08577403
; Patent No. 5775693
; GENERAL INFORMATION:
; APPLICANT: Burgess, Diane G.
; APPLICANT: Dooner, Hugo K.

TITLE OF INVENTION: Pea ADP-glucose pyrophosphorylase Subunit
TITLE OF INVENTION: Genes and Their Uses
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend Kourie and Crew
STREET: One Market Plaza, Steuart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/577,403
FILING DATE: 22-DEC-1995
CLASSIFICATION: 800
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: 08/097,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 12176-35
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2422
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-577-403-4

Query Match 3.1%; Score 97; DB 1; Length 516;
Best Local Similarity 20.9%; Pred. No. 0.32;
Matches 123; Conservative 69; Mismatches 172; Indels 224; Gaps 32;

QY 1 MASFTATAAVSGRWLGNGHTOPPLSSSGSSDLSYCSLPMASRYTRKLNVSAA-----L 54
DB 1 MASMAAIGYK-----VPPSSSSSS-----SSSSSKAIARNLSFTSSQLGDKXI 46
QY 55 HRPPLAHFPKQSSNSPAIVYKPKA-KESNTKQKNLFQRAAAALDAABEFLV---SHEKL 110
DB 47 F---VSGTRSSGRRNPFLVSPRAVSDSKNSQTCIDPDASRYL---GILGGGAGTR 99
QY 111 HPL-PTADPSVQIAGNFAPVNEQPRRMIPVVGKLPDSIKGVYVNGANPLHEPYTGHH 169
DB 100 YPLTKRKRAKPAVPLGANYLI-----DIPVSNCLNSISIKIYVLTQTN---SASLNH 149
QY 170 F-----FDGGMVAHVAKFE-----HGSA---SYACFTQTN----- 197
DB 150 LSRAYASNLGKYKNEGFEVLAQQSPENPNWFOGTADAVROQLMFEHNVLLEYVLVG 209
QY 198 ----RVOQROGLGRVFPRAIGELHGHGTIARLMFLYANAAGIYVPAIGTGAANAGLY 253
DB 210 DHLRYRDYR-----FIQAHRESADITVASLPMDEARATA-----FGLMK 250
QY 254 FN--GRLLAMSE-----DD-----LPYQVQTPNG-----D 277
DB 251 IDBEGRIVFSEKPKGEQIKAMKVDTTILGLDDERAKEMPY---IASMGITYVYSKHVMD 307
QY 278 L--KTGVGFDEGQLESTMIAPKVDPESEL---FALSYD----- 313
DB 308 LRDKRPKANDPGE-----VIRGATELGRLVQYLVLDGVYEDIGTIEAFYANVL 357
QY 314 -VYSKP---YLKIFRSPDGTSKSPDVEIQLDQPTMHDAITENY---VPPDOYV--- 362
DB 358 GITKKRPVDFSPFSSPIYTOP-----RYLPPSKMLDADITDSVIGECVINKCKTIHNS 412

QY 363 VEKLPEMI-----RGSSPVYDKNKVAREGILDKYAE 394
DB 413 VVGLRSCISEGAIIEDTLTMGADYETDADRFRFLAAKGVPGLGKNSIKRAIITDKNAR 472
QY 395 DSSNIRKIDAPDPCFHLWNAAEPE---TDEVVIGSCMTPPDSI 437
DB 473 IGDDVYKILNSD-----NVOEAAARETEGYFKSGIYVIVIKDALIPSGTV 515

RESULT 13
US-09-335-409-3
; Sequence 3, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-3

Query Match 3.0%; Score 96; DB 3; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.4;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;

QY 10 VSGRWLGNGHTOPPLSSSQS-SDLSY-----CSSLPMASRYTRK-----LNVSALHT 56
DB 645 VSEAGYEGSGDQPPMMPITQPSDLAVITYTSSGTGLPKGVMTDHRGAVVTIIDINERFET 704
QY 57 PALHFPKQSSNS-----PAIVYKPKA-KESNTKQKNLFQRAAA---AL 97
DB 705 GPGDRVLALSISFELSYDVFEGILAAAGTIYVPAOSKLRDPAHNAELIEREKVYWNVS 764
QY 98 DAAEGFLVSHKLNH--LPKTDAPSVQIAGNFAPV---NEQVVRNLPVY---GKLPS 148
DB 765 PALMRVLVHEFGRPDSIARSLRLSL-LSGDMIPVGLPGLQAIRGVSVISLGGATEAS 823
QY 149 I--KGVYVNGANPLHEPYTGHHFFDGDGMVAHVAKFEHGSASVACRF--TQNRFOERQL 205
DB 824 IMSIGYPVN-----VDLSMASIYGRPLRQGTHTVLDLEALE 860
QY 206 GRVPEPKAIGELH-GHTGIARLMLEYAR-----AAAGIYDPAHGTGVANAGLVYENGRL 259
DB 861 PRPVW--VPGQLYIGGVGLA---LGWRODEKTRKSFVLHPETGERLYXTGDL---GRYL 912
QY 260 AMSEDDLPYQVQITPBGDKTYGRFED-----GQLSTIMAHKXVDPESGELFA 309
DB 913 -----PDGNTPEFGREDNQIKLRGYRVELGEIEETLSHPNV--RDAVIYR 956
QY 310 LSYDVYSKPYLKYFRSPDGTSKSPDVEIQLDQPTMHDA---ITENFVYVPPDOYVFKLP 367
DB 957 VGNDAAANKLLIY--VPEGTRRAAEQDASIKTERITDRRAHALEADGLSDGERVQFKLA 1014
QY 368 EM-IR---GSPVYVYDKNKVAREFGILDKYAEDSSNIKWIDAP 405
DB 1015 RHGIRRDLDGKRPVVDLTGQDPREAGLDVYARRRSVATFLEAP 1056

RESULT 14
US-09-568-102-3

```
; Sequence 3, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-102-3
```

```
Query Match          3.0%; Score 96; DB 4; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.4;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;

QY 10 VSGRMLGNNHTQPPPLSSQS--SDLSY-----CSSLPMASTRYTRK-----LNVSALHT 56
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 645 VSEAGVEGDGQPPMPPIQTSPDLAVIYITSGSTGLPKGVMIDHGA VNTIIDINERFEI 704

QY 57 PPALHPKQSSNS-----PAIVKPKAK--ESNTKQMLFORAAA---AL 97
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 705 GPGRVNLALSLSPDLSDYDFGILAAAGTIYVPDASKLRDPAHMAELIEREKVTWMSV 764

QY 98 DAEEGFLVSHKELHP--LPKTADPSVQIAGNFAPV---NEOPVRNLPVY---GKLPS 148
   || | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 765 PALMRLVHEHFGRPDSLARSRLSL--LSGDMIPVGLPGLDQAIKPGVSVISLGATEAS 823

QY 149 I--KGVYVRNGANPLHEPVYTGHHFFDGDGMVAHVFEHGSASYACRF--TQTRRFYQEROL 205
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 824 IWSIGYPRVN-----VDLSMASIPIYGRPLRNQTFHVIDEAL 860

QY 206 GRPVPKAIIGELH--GHTGIARLMLFYAR-----AAAGIYDPAHGTGVANAGLVYENGRL 259
   || | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 861 PRPVW--VPGQLYIGGVGLA---LGWDEDEKTRKSFVHPETGERLYKTGDL---GRTL 912

QY 260 AMSEDDLRYQVOITPNGDLKTYGREFD-----GQLESTMIAPHKVDPESGELFA 309
   || | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 913 -----PDGNIIEFWGREDNQIKLGYRVLEGEIETILKSHPNV--RDAVIY 956

QY 310 LSYDVVSKPYLYKFFRSPDGTSPDVEIQLDQPTMMHFA--ITENFVVVPDQOVVFKLP 367
   : | : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 957 VGNDAANKLLAY--VVEGTRRRAAEODASLTERIDARAHAAEDGLSDGERVQFKLA 1014

QY 368 EM-IR---GSPVYVDKNKVAFAFGILDKYAEDSSNIKWIDAP 405
   : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1015 RHGLRRDLDGKRPVVDLTGQDPREAGLDVYARRRSVTFLEAP 1056
```

```
RESULT 15
US-09-567-969-3
; Sequence 3, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
```

```
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
```

```
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1410
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-567-969-3
```

```
Query Match          3.0%; Score 96; DB 4; Length 1410;
Best Local Similarity 23.2%; Pred. No. 2.4;
Matches 107; Conservative 57; Mismatches 182; Indels 116; Gaps 25;

QY 10 VSGRMLGNNHTQPPPLSSQS--SDLSY-----CSSLPMASTRYTRK-----LNVSALHT 56
   || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 645 VSEAGVEGDGQPPMPPIQTSPDLAVIYITSGSTGLPKGVMIDHGA VNTIIDINERFEI 704

QY 57 PPALHPKQSSNS-----PAIVKPKAK--ESNTKQMLFORAAA---AL 97
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 705 GPGRVNLALSLSPDLSDYDFGILAAAGTIYVPDASKLRDPAHMAELIEREKVTWMSV 764

QY 98 DAEEGFLVSHKELHP--LPKTADPSVQIAGNFAPV---NEOPVRNLPVY---GKLPS 148
   || | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 765 PALMRLVHEHFGRPDSLARSRLSL--LSGDMIPVGLPGLDQAIKPGVSVISLGATEAS 823

QY 149 I--KGVYVRNGANPLHEPVYTGHHFFDGDGMVAHVFEHGSASYACRF--TQTRRFYQEROL 205
   || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 824 IWSIGYPRVN-----VDLSMASIPIYGRPLRNQTFHVIDEAL 860

QY 206 GRPVPKAIIGELH--GHTGIARLMLFYAR-----AAAGIYDPAHGTGVANAGLVYENGRL 259
   || | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 861 PRPVW--VPGQLYIGGVGLA---LGWDEDEKTRKSFVHPETGERLYKTGDL---GRTL 912

QY 260 AMSEDDLRYQVOITPNGDLKTYGREFD-----GQLESTMIAPHKVDPESGELFA 309
   || | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 913 -----PDGNIIEFWGREDNQIKLGYRVLEGEIETILKSHPNV--RDAVIY 956

QY 310 LSYDVVSKPYLYKFFRSPDGTSPDVEIQLDQPTMMHFA--ITENFVVVPDQOVVFKLP 367
   : | : | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 957 VGNDAANKLLAY--VVEGTRRRAAEODASLTERIDARAHAAEDGLSDGERVQFKLA 1014

QY 368 EM-IR---GSPVYVDKNKVAFAFGILDKYAEDSSNIKWIDAP 405
   : | | | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1015 RHGLRRDLDGKRPVVDLTGQDPREAGLDVYARRRSVTFLEAP 1056
```

Search completed: July 19, 2002, 10:12:35
Job time: 92 sec

